

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:21:39 ; Search time 41.4359 Seconds  
(without alignments)  
109.103 Million cell updates/sec

Title: US-10-070-566-6

Perfect score: 83

Sequence: 1 GVTAPDTRPAPGSTA 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83	100.0	16	2 AAR84861	Aar84861 MUC-1 epi
2	83	100.0	16	2 AAW35740	Aaw35740 Keyhole 1
3	83	100.0	16	2 AAW01775	Aaw01775 Human MUC
4	83	100.0	16	2 AAW67591	Aaw67591 T-cell ac
5	83	100.0	16	2 AAY09172	Aay09172 MUC-1 muc
6	83	100.0	16	4 AAB73352	Aab73352 Human pol
7	83	100.0	16	4 AAU00488	Aau00488 Human muc
8	83	100.0	16	4 AAB97223	Aab97223 Mucin pep
9	83	100.0	16	6 AAE33942	Aae33942 MUC 1 pep
10	83	100.0	17	6 AAE33940	Aae33940 MUC 1 pep
11	83	100.0	18	4 AAU07428	Aau07428 Mucin MUC
12	83	100.0	18	6 AAE33945	Aae33945 MUC 1 pep
13	83	100.0	19	4 AAB83309	Aab83309 Lipopepti
14	83	100.0	19	4 AAB83310	Aab83310 Lipopepti
15	83	100.0	19	4 AAB83308	Aab83308 Lipopepti
16	83	100.0	19	4 AAU07427	Aau07427 Mucin MUC
17	83	100.0	19	4 AAU07431	Aau07431 Mucin MUC
18	83	100.0	19	6 AAE33944	Aae33944 MUC 1 pep
19	83	100.0	20	2 AAR77000	Aar77000 Human muc
20	83	100.0	20	2 AAW21976	Aaw21976 Human MUC
21	83	100.0	20	2 AAW22250	Aaw22250 Muc pep 2
22	83	100.0	20	2 AAW27393	Aaw27393 Human epi
23	83	100.0	20	2 AAY25112	Aay25112 Synthetic
24	83	100.0	20	2 AAW67589	Aaw67589 T-cell ac
25	83	100.0	20	3 AAY80111	Aay80111 MUC-1 rep

26	83	100.0	20	3 AAB09915	Aab09915 MUC-1 der
27	83	100.0	20	3 AAB29928	Aab29928 Cytochemi
28	83	100.0	20	4 AAE12708	Aae12708 Human MUC
29	83	100.0	20	4 AAU04015	Aau04015 Mucin, MU
30	83	100.0	20	5 AAU82063	Aau82063 Antigenic
31	83	100.0	20	5 ABB08278	Abb08278 Human can
32	83	100.0	20	6 ABUS8800	Abu58800 Mucin 1 (
33	83	100.0	20	6 AAE33943	Aae33943 MUC 1 pep
34	83	100.0	20	6 AAE33957	Aae33957 Dilipidat
35	83	100.0	20	6 AAE30200	Aae30200 Human MUC
36	83	100.0	20	7 ADB84183	Adb84183 Human MUC
37	83	100.0	20	7 ADD88878	Add88878 Human MUC
38	83	100.0	21	2 AAW21975	Aaw21975 Human MUC
39	83	100.0	21	2 AAW22249	Aaw22249 Muc pep 1
40	83	100.0	21	2 AAY25111	Aay25111 Synthetic
41	83	100.0	21	3 AAY96171	Aay96171 MUC1 repe
42	83	100.0	21	3 AAY96173	Aay96173 MUC1 repe
43	83	100.0	21	3 AAY96169	Aay96169 MUC1 repe
44	83	100.0	21	3 AAY96180	Aay96180 MUC1 repe
45	83	100.0	21	3 AAY96179	Aay96179 MUC1 repe

ALIGNMENTS

RESULT 1  
AAR84861  
ID AAR84861 standard; peptide; 16 AA.  
XX  
AC AAR84861;  
XX  
DT 30-MAY-1996 (first entry)  
XX  
DE MUC-1 epitope.  
XX  
KW Antigen; epitope; cell mediated; immune specific; cancer; infection;  
KW infestation; mucin-1; MUC-1; tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO9527505-A1.  
XX  
PD 19-OCT-1995.  
XX  
PF 12-APR-1995; 95WO-US004540.  
XX  
PR 12-APR-1994; 94US-00229606.  
XX  
PA (BIOM-) BIOMIRA INC.  
XX  
PI Longenecker BM, Ding L, Reddish MA, Koganty RR;  
XX  
DR WPI; 1995-373528/48.  
XX  
PT New cell-mediated immune-specific immunogenic compsn. - used in  
PT prophylaxis and treatment of cancer, microbial infections, viral  
PT infections and parasitic infestations.  
XX  
PS Example 1; Page 62; 141pp; English.  
XX  
CC A claimed cell-mediated immune (CMI)-specific immunogenic compsn.  
CC comprises a mixt. of primary antigen bearing a primary epitope, i.e. the  
CC present peptide, with an immunomodulatory peptide (IP). The IP comprises  
CC an allopetide moiety of at least 5 amino acids, whose sequence  
CC corresponds to a polymorphic region of a MHC encoded polymorphic Class I  
CC or II antigen. The compsn. can be used to elicit a CMI-specific response  
CC which is prophylactic, or therapeutic for, e.g. microbial and viral  
CC infections, parasitic infestations and cancer, partic. MUC-1 expressing  
XX  
SQ tumour cells when the present peptide is the primary epitope  
Sequence 16 AA;  
Query Match 100.0%; Score 83; DB 2; Length 16;

```
Best Local Similarity 100.0%; Pred. No. 4.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16
   |||||
Db 1 GVTSAPDTRPAPGSTA 16

RESULT 2
AAW35740
ID AAW35740 standard; peptide; 16 AA.
XX
AC AAW35740;
XX
DT 02-APR-1998 (first entry)
XX
DE Keyhole limpet haemocyanin BP-1-7-KLH.
XX
KW MUC1; mucin; vaccine; keyhole limpet haemocyanin; KHL; cancer;
KW immunogenic protein; immune response.
XX
OS Synthetic.
XX
PN WO9734921-A1.
XX
PD 25-SEP-1997.
XX
PF 20-MAR-1997; 97WO-US004493.
XX
PR 20-MAR-1996; 96US-0013775P.
XX
PA (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI Livingston PO, Zhang S;
XX
DR WPI; 1997-480158/44.
XX
PT Vaccine effective against cancer of the breast, prostate, colon, lung or
PT pancreas - comprising mucin peptide, especially MUC1, conjugated to
PT immunogenic protein, especially keyhole limpet haemocyanin.
XX
PS Disclosure; Page 24; 45pp; English.
XX
CC This keyhole limpet haemocyanin (KLH) BP-1-7-KLH is an immunogenic
CC protein that can stimulate or enhance immune response in a subject. It is
CC used in a vaccine capable of producing an immune response which
CC recognises a mucin. The mucin peptide is selected from MUC1 peptide
CC group. The vaccine comprises an amount of the mucin peptide conjugated to
CC this immunogenic protein KLH, together with an adjuvant and a vehicle.
CC The vaccine can be used to induce an immune response in patients
CC suffering from a cancer of the type where the cancer cells have mucin on
CC their surface, e.g. breast cancer, prostate cancer, lung cancer, colon
CC cancer or pancreas cancer
XX
SQ Sequence 16 AA;

Query Match 100.0%; Score 83; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16
   |||||
Db 1 GVTSAPDTRPAPGSTA 16

RESULT 3
AAW01775
ID AAW01775 standard; peptide; 16 AA.
XX
AC AAW01775;
XX
DT 20-AUG-1997 (first entry)
XX
```

```
Human MUC1 antigen.
T cell antigen epitope; Th1 specific immunomodulator; ovarian cancer;
tuberculosis; malaria; breast cancer; pancreatic cancer;
respiratory syncytial virus infection; leishmaniasis; leprosy;
candidiasis.
Homo sapiens.
WO9640066-A1.
19-DEC-1996.
07-JUN-1996; 96WO-US009951.
07-JUN-1995; 95US-00480499.
(UYAL-) UNIV ALBERTA.
Samuel J, Kwon GS;
WPI; 1997-051316/05.
Compsn. contg. T cell antigen epitope and Th1 specific immuno:modulator -
useful for eliciting Th1 specific immune response to treat e.g. cancer,
tuberculosis, malaria, etc.
Claim 13; Page 53; 79pp; English.
A novel composition is able to elicit Th1 type immune responses against
short T-cell epitope-containing peptides without the use of traditional
carrier proteins and immunoadjuvants. The composition involves: a slow
release vehicle; an immunogenically effective amount of a synthetic
peptide consisting of an 11-14 amino acid sequence containing at least
one T cell antigen epitope; and an immunogenically effective amount of a
Th1 specific immunomodulator. The present sequence is human MUC1 antigen,
which is a specifically claimed example of an amino acid sequence used in
the new composition. The composition can be used to elicit a Th1 specific
immune response, preferably a protective immune response, to treat e.g.
breast, pancreatic and ovarian cancer, respiratory syncytial virus
infections, leishmaniasis, malaria, tuberculosis, leprosy and candidiasis
Sequence 16 AA;

Query Match 100.0%; Score 83; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16
   |||||
Db 1 GVTSAPDTRPAPGSTA 16

RESULT 4
AAW67591
ID AAW67591 standard; peptide; 16 AA.
XX
AC AAW67591;
XX
DT 02-MAR-1999 (first entry)
XX
DE T-cell activation peptide #9.
XX
KW Activated T helper cell; CD4+; cytotoxic T cell; CD8+; liposome; epitope;
KW peripheral blood lymphocyte; antigen-presenting cell; APC; virus; tumour;
KW bacterium; parasite; cytokine; vaccine; cancer; malaria; HIV; hepatitis;
KW tuberculosis.
XX
OS Synthetic.
XX
PN WO9850527-A1.
XX
PD 12-NOV-1998.
```



XX 07-MAY-1998; 98WC-US009288.  
 XX 08-MAY-1997; 97US-0045949P.  
 XX (BIOM-) BIOMIRA INC.  
 XX Agrawal B, Krantz MJ, Reddish MA, Longenecker BM;  
 XX WPI; 1999-034715/03.  
 XX Method of activation of T cells - by exposure to antigen-presenting cells  
 XX loaded with antigen in liposome, used for, e.g. treating cancer and  
 XX microbial infections.  
 XX Claim 14; Page 49; 75pp; English.  
 XX Peptides AAW67583-W67611 are used to produce activated T helper (CD4+) and cytotoxic (CD8+) T-cells. The activated T cells are produced by treating peripheral blood lymphocytes with liposome-encapsulated peptide antigen to generate Ag-loaded antigen-presenting cells (APC), contacting naive or anergic T-cells with these APC, and isolating the resulting activated T-cells. The cells are specific for a particular antigen, particularly one derived from a tumour, but also those from viruses, bacteria and other parasites. It can also be used to identify antigens and epitopes able to generate an Ag-specific T-cell response (by assessing proliferation and cytokine release). Also the Ag-loaded APC can be used as cellular vaccines for treating cancer (claimed) or other diseases (e.g. malaria, human immune deficiency virus infection, hepatitis, tuberculosis). The activated T-cells can be used to treat the same conditions by adoptive T-cell transfer therapy  
 XX Sequence 16 AA;  
 XX Query Match 100.0%; Score 83; DB 2; Length 16;  
 XX Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 Db |||||  
 1 GVTSAPDTRPAPGSTA 16  
 RESULT 5  
 AAY09172  
 ID AAY09172 standard; peptide; 16 AA.  
 XX AAY09172;  
 XX 19-JUL-1999 (first entry)  
 XX MUC-1 mucin peptide derivative.  
 XX MUC-1; mucin; cytokine; immunosuppression; anergy; adenocarcinoma; breast; colon; lung; ovarian; pancreatic; cancer; under-glycosylated; immunotherapy.  
 XX Synthetic.  
 XX Homo sapiens.  
 XX WO9923114-A1.  
 XX 14-MAY-1999.  
 XX 30-OCT-1998; 98WO-US022644.  
 XX 31-OCT-1997; 97US-0064146P.  
 XX 12-NOV-1997; 97US-0065209P.  
 XX (BIOM-) BIOMIRA INC.  
 XX Agrawal B, Reddish MA, Longenecker BM;  
 XX

DR WPI; 1999-313304/26.  
 XX MUC-1 mucin derivatives useful in cancer therapy.  
 XX Claim 2; Page 15; 25pp; English.  
 XX The invention relates to a MUC-1 mucin derivative consisting of a single MUC-1 core repeat. The MUC-1 derivative, optionally linked to a cytokine, is useful for the alleviation of immunosuppression or anergy. This is useful for treatment of human adenocarcinomas, such as breast, colon, lung, ovarian and pancreatic cancers that abundantly over express and secrete under-glycosylated MUC-1 protein. Addition of a cytokine supports the immunotherapy of the MUC-1 derivative. The present sequence represents a claimed example of a MUC-1 derivative  
 XX Sequence 16 AA;  
 XX Query Match 100.0%; Score 83; DB 2; Length 16;  
 XX Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 Db |||||  
 1 GVTSAPDTRPAPGSTA 16  
 RESULT 6  
 AAB73352  
 ID AAB73352 standard; peptide; 16 AA.  
 XX AAB73352;  
 XX 22-MAY-2001 (first entry)  
 XX Human polymorphic epithelial mucin (PEM1p) epitope.  
 XX Oral vaccine composition; oral adjuvant; mucosal adjuvant; food grade saponin; Quillaja saponaria; mucosal immune response; systemic immune response; gastrointestinal infection; urogenital infection; hepatitis B virus surface antigen; HBsAg; human polymorphic epithelial mucin epitope; PEM1p.  
 XX Homo sapiens.  
 XX WO2000117555-A2.  
 XX 15-MAR-2001.  
 XX 11-SEP-2000; 2000WO-GB003492.  
 XX 09-SEP-1999; 99GB-00021347.  
 XX (DOWC ) DOW CHEM CO.  
 XX Brennan FR, Hamilton WDO;  
 XX WPI; 2001-235162/24.  
 XX New oral vaccines comprising an antigen and a food grade saponin as adjuvant, useful inducing mucosal and systemic immune responses, and as prophylaxis against gastrointestinal and urogenital infections or disease agents.  
 XX Example 7; Page 11; 16pp; English.  
 XX The invention relates to a novel oral vaccine composition comprising an antigen and a food grade saponin. The food grade saponin acts as a mucosal adjuvant, especially as an oral adjuvant and is preferably derived from extracts from the bark of Quillaja species trees (e.g., the Chilean soap bark tree, Quillaja saponaria). Compositions of the invention are useful as oral vaccines for prophylactic or therapeutic use. Such vaccines are useful for the induction of mucosal and systemic immune responses, and for prophylaxis directed against gastrointestinal

CC and urogenital infections or disease agents, as well as any pathogens  
CC whose biology necessitates interaction with any mammalian mucosal  
CC surface. Food grade saponins are used to specifically boost the immune  
CC response to an antigen which has been used to prime the immune system of  
CC an individual. This is particularly useful for boosting a primed but  
CC critically low level immune reaction to hepatitis B virus surface antigen  
CC (HBsAg). Compared with the highly purified Quillaja extracts previously  
CC used as oral adjuvants (i.e., Quila and QS21), food grade saponins are  
CC cheaper. However, as they are of a higher purity compared with crude  
CC Quillaja bark extracts, they avoid the toxicity problems associated with  
CC such crude extracts, and may be introduced into the human food chain as  
CC they are considered safe for human ingestion. The present sequence  
CC represents a human polymorphic epithelial mucin (PEM1p) epitope which was  
CC coupled to a carrier protein and orally administered to mice either alone  
CC or in combination with food grade saponins

XX  
SQ Sequence 16 AA;  
Query Match 100.0%; Score 83; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GVTSAPDTRPARGSTA 16  
Db 1 GVTSAPDTRPARGSTA 16  
|||||

RESULT 7  
AAU00488  
ID AAU00488 standard; peptide; 16 AA.  
XX  
AC AAU00488;  
XX  
DT 04-JUL-2001 (first entry)  
XX  
DE Human mucin peptide MUC1(16).  
XX  
KW Human; polymorphic epithelial cell mucin; PEM; chimeric virus particle;  
KW CVP; plant virus coat protein; comovirus; CPWV; cowpea mosaic virus;  
KW SBMV; Southern bean mosaic virus; LTSV; red clover necrotic mosaic virus;  
KW RCNMV; MUC1; tumour; cancer vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200118199-A1.  
XX  
PD 15-MAR-2001.  
XX  
PF 11-SEP-2000; 2000WO-GB003500.  
XX  
PR 09-SEP-1999; 99GB-00021337.  
XX  
PA (DWC ) DOW CHEM CO.  
XX  
PI Bendig MM, Jones TD, Longstaff M, Hellendoorn K;  
XX  
DR WPI; 2001-244570/25.  
DR N-ESDB; AAS01380.  
XX  
XX  
PT New chimeric plant virus particles with an immunogenically active peptide  
PT of a tumor-associated mucin, useful as a vaccine or for the manufacture  
PT of a vaccine for treating and/or preventing tumors and/or cancer.  
XX  
PS Claim 7; Fig 8B; 63pp; English.  
XX  
CC The present sequence representing human mucin peptide MUC1(16) is 1 of 7  
CC polymorphic epithelial cell mucin (PEM) peptides (AAU00483-AAU00489). A  
CC novel chimeric virus particle (CVP) is generated from a plant virus coat  
CC protein with a beta barrel structure and modified by insertion of an  
CC immunogenically active mucin peptide at an immunogenic site in the coat  
CC protein. The mucin peptide epitopes are preferably inserted into the coat  
CC protein of a plant virus such as the comovirus CPWV (cowpea mosaic  
CC virus). The ability of the chimeric virus particle CPWV-MUC1(16) to

CC elicit antibodies, which can cause regression of tumours expressing the  
CC MUC1 protein, is demonstrated in a mouse tumour model. Other examples of  
CC CVPs include the insertion of the mucin MUC1(16) peptide into the coat  
CC proteins of other plant viruses such as SBMV (Southern bean mosaic virus;  
CC AAU00490), LTSV (AAU00491) and RCNMV (red clover necrotic mosaic virus;  
CC AAU00492). The CVP is useful as or as part of a vaccine particularly for  
CC treating and preventing tumours and cancer. The CVP provides advantages  
CC over prior art antigen-presenting means since conventional live animal  
CC virus vectors can be avoided; as can the need for separate mucin peptide  
CC synthesis and chemical-coupling to a conventional carrier. Also, the CVP  
CC is shown to induce good mucosal immunity, and does not require the  
CC addition of exogenous adjuvants to induce a strong immune response

XX  
SQ Sequence 16 AA;  
Query Match 100.0%; Score 83; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPARGSTA 16  
Db 1 GVTSAPDTRPARGSTA 16  
|||||

RESULT 8  
AAB97223  
ID AAB97223 standard; peptide; 16 AA.  
XX  
AC AAB97223;  
XX  
DT 19-JUL-2001 (first entry)  
XX  
DE Mucin peptide SEQ ID 37.  
XX  
KW Viral capsid; immunostimulant; cytostatic; antibacterial; antiviral;  
KW antifungal; vaccine; cancer; CPWV; mucin.  
XX  
OS Unidentified.  
XX  
PN WO200127282-A1.  
XX  
PD 19-APR-2001.  
XX  
PF 13-OCT-2000; 2000WO-US028430.  
XX  
PR 14-OCT-1999; 99GB-00024352.  
XX  
PA (DWC ) DOW CHEM CO.  
XX  
PI Hellendoorn K, Jones T;  
XX  
DR WPI; 2001-282032/29.  
XX  
PT New chimeric viral particle capable of assembly in a host cell or tissue  
PT useful as a vaccine for cancer or protection against disease caused by  
PT pathogens.  
XX  
PS Example 11; Page 55; 88pp; English.  
XX  
CC This invention relates to chimeric viral particles, in which the internal  
CC side of the viral capsid has at least one exogenous peptide attached. The  
CC invention includes methods for the production of the chimeric viral  
CC particles in which exogenous epitopes are expressed on the interior of  
CC the viral capsid. The chimeric viral particles can be used to produce  
CC immunostimulant, cytostatic, antibacterial, antiviral and antifungal  
CC activity. The viral particles are used as vaccines to induce an  
CC immunogenic or an antigenic response in humans or animals for the  
CC protection from diseases caused by pathogens. The viral particles may  
CC also be used in the treatment of cancer. Vectors encoding modified Cowpea  
CC mosaic virus (CPWV) genomes containing peptide inserts in the coat  
CC proteins are used in the production of the chimeric viral particles. The  
CC present sequence represents a mucin derived epitope. The peptide is used  
CC to test the efficacy of combining epitopes internally with epitopes

CC externally on a single chimeric virus particle

SQ Sequence 16 AA;

Query Match 100.0%; Score 83; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16

Db 1 GVTSAPDTRPAPGSTA 16

RESULT 9

AAE33942

ID AAE33942 standard; peptide; 16 AA.

XX

AC AAE33942;

DT 02-MAY-2003 (first entry)

DE MUC 1 peptide #4.

XX Liposome; vaccine; immune response; MUC 1 lipopeptide; immunomodulator;

KW T cell proliferation; antibody production.

XX Unidentified.

OS WO200276485-A2.

XX 03-OCT-2002.

XX 27-MAR-2002; 2002WO-IB002188.

XX 27-MAR-2001; 2001US-0278698P.

XX (BUDZ/) BUDZYNSKI W A.

XX Budzynski WA, Koganty RR, Krantz MJ, Longenecker MB;

XX WPI; 2003-046750/04.

XX New liposome vaccines comprising at least one monolipopeptide and at

PT least one dilipopeptide, useful for modulating the immune response in

PT vivo, particularly humoral and cellular immune responses.

XX Disclosure; Page 4; 5lpp; English.

XX The present invention relates to liposomal compositions comprising at

CC least one liposome that comprises at least one monolipopeptide and at

CC least one dilipopeptide derived from a protein associated with a disease

CC selected from the group consisting of tuberculosis, malaria, cancer and

CC hepatitis B. The monolipopeptide or dilipopeptide is designed from MUC 1

CC protein. The composition is useful as a vaccine for modulating the immune

CC response to the peptide in vivo, particularly humoral and cellular immune

CC responses where the relative amounts of monolipopeptide and dilipopeptide

CC modulate the relative intensities of T cell proliferation and antibody

CC production. The present sequence is MUC 1 peptide

XX Sequence 16 AA;

SQ

Query Match 100.0%; Score 83; DB 6; Length 16;

Best Local Similarity 100.0%; Pred. No. 4.6e-05;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16

Db 1 GVTSAPDTRPAPGSTA 16

RESULT 10

AAE33940

ID AAE33940 standard; peptide; 17 AA.

XX

AC AAE33940;

DT 02-MAY-2003 (first entry)

XX MUC 1 peptide #2.

DE Liposome; vaccine; immune response; MUC 1 lipopeptide; immunomodulator;

KW T cell proliferation; antibody production.

XX Unidentified.

OS WO200276485-A2.

XX 03-OCT-2002.

XX 27-MAR-2002; 2002WO-IB002188.

XX 27-MAR-2001; 2001US-0278698P.

XX (BUDZ/) BUDZYNSKI W A.

XX Budzynski WA, Koganty RR, Krantz MJ, Longenecker MB;

XX WPI; 2003-046750/04.

XX New liposome vaccines comprising at least one monolipopeptide and at

PT least one dilipopeptide, useful for modulating the immune response in

PT vivo, particularly humoral and cellular immune responses.

XX Claim 51; Page 44; 5lpp; English.

XX The present invention relates to liposomal compositions comprising at

CC least one liposome that comprises at least one monolipopeptide and at

CC least one dilipopeptide derived from a protein associated with a disease

CC selected from the group consisting of tuberculosis, malaria, cancer and

CC hepatitis B. The monolipopeptide or dilipopeptide is designed from MUC 1

CC protein. The composition is useful as a vaccine for modulating the immune

CC response to the peptide in vivo, particularly humoral and cellular immune

CC responses where the relative amounts of monolipopeptide and dilipopeptide

CC modulate the relative intensities of T cell proliferation and antibody

CC production. The present sequence is MUC 1 peptide

XX Sequence 17 AA;

SQ

Query Match 100.0%; Score 83; DB 6; Length 17;

Best Local Similarity 100.0%; Pred. No. 4.9e-05;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16

Db 1 GVTSAPDTRPAPGSTA 17

RESULT 11

AAU07428

ID AAU07428 standard; peptide; 18 AA.

XX AAU07428;

XX 18-DEC-2001 (first entry)

XX Mucin MUC-1-derived peptide, BP1-228.

XX MUC-1; mucin; cytostatic; virucide; vaccine; adjuvant peptide; antigen;

KW immune response; cancer; viral disease.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 17

FT /note= "Ser-myristyl"

XX

PN WO200170265-A2.  
 XX 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-IB000703.  
 XX 24-MAR-2000; 2000US-0191736P.  
 PR (BIOM-) BIOMIRA INC.  
 XX Agrawal B, Longenecker MB, Parker J;  
 PI WPI; 2001-611447/70.  
 XX Novel vaccine for treating and preventing disorders such as cancer and  
 PT viral diseases, comprises mucin MUC-1-based adjuvant peptide and an  
 PT antigen.  
 XX Claim 6; Page 6; 22pp; English.  
 PS The invention relates to a vaccine composition comprising a mucin MUC-1-  
 CC based adjuvant peptide and an antigen. The vaccine composition is useful  
 CC for stimulating the immune response of a patient, by administering the  
 CC vaccine. Alternatively, this can be done by contacting ex vivo a T-cell  
 CC from the patient with the vaccine and administering the contacted cells  
 CC to the patient which stimulates the immune system of the patient. The  
 CC vaccine is useful for treating and preventing disorders such as cancer  
 CC and viral diseases. The vaccine is effective in generating an immune  
 CC response to an antigen against which the patient does not respond. The  
 CC present sequence represents the amino acid sequence of mucin MUC-1-  
 CC derived peptide, BP1-228, which was used in the vaccine composition  
 XX  
 SQ Sequence 18 AA;  
 Query Match 100.0%; Score 83; DB 4; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSA PDTRPAPGSTA 16  
 Db 1 GVTSA PDTRPAPGSTA 16  
 RESULT 12  
 AAEE33945  
 ID AAEE33945 standard; peptide; 18 AA.  
 XX  
 AC AAEE33945;  
 XX  
 DT 02-MAY-2003 (first entry)  
 XX  
 DE MUC 1 peptide, BP1-228.  
 XX  
 KW Liposome; vaccine; immune response; MUC 1 lipopeptide; immunomodulator;  
 KW T cell proliferation; antibody production.  
 XX Unidentified.  
 OS  
 XX Key Location/Qualifiers  
 FH Modified-site 17  
 FT /note= "Linked to myristyl group"  
 FT  
 XX WO200276485-A2.  
 PN  
 XX 03-OCT-2002.  
 PD  
 XX 27-MAR-2002; 2002WO-IB002188.  
 PF  
 XX 27-MAR-2001; 2001US-0278698P.  
 PR  
 XX (BUDZ/) BUDZYNSKI W A.  
 PA  
 XX Budzynski WA, Koganty RR, Krantz MJ, Longenecker MB;

XX WPI; 2003-046750/04.  
 DR  
 XX New liposome vaccines comprising at least one monolipopeptide and at  
 PT least one dilipopeptide, useful for modulating the immune response in  
 PT vivo, particularly humoral and cellular immune responses.  
 XX  
 PS Example 2; Page 26; 51pp; English.  
 XX  
 CC The present invention relates to liposomal compositions comprising at  
 CC least one liposome that comprises at least one monolipopeptide and at  
 CC least one dilipopeptide derived from a protein associated with a disease  
 CC selected from the group consisting of tuberculosis, malaria, cancer and  
 CC hepatitis B. The monolipopeptide or dilipopeptide is designed from MUC 1  
 CC protein. The composition is useful as a vaccine for modulating the immune  
 CC response to the peptide in vivo, particularly humoral and cellular immune  
 CC responses where the relative amounts of monolipopeptide and dilipopeptide  
 CC modulate the relative intensities of T cell proliferation and antibody  
 CC production. The present sequence is MUC 1 peptide  
 XX  
 SQ Sequence 18 AA;  
 Query Match 100.0%; Score 83; DB 6; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSA PDTRPAPGSTA 16  
 Db 1 GVTSA PDTRPAPGSTA 16  
 RESULT 13  
 AAB83309  
 ID AAB83309 standard; peptide; 19 AA.  
 XX  
 AC AAB83309;  
 XX  
 DT 18-SEP-2001 (first entry)  
 XX  
 DE Lipopeptide BP1-217.  
 XX  
 KW Lipopeptide; Lipid A; MUC1; mucin; analogue; immune response; vaccine;  
 KW anti-tumour agent; antibiotic; cancer; therapy.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 17 /note= "S(Lipo)"  
 FT Modified-site 18 /note= "S(Lipo)"  
 FT  
 XX WO200136433-A2.  
 PN  
 XX 25-MAY-2001.  
 PD  
 XX 15-NOV-2000; 2000WO-US031281.  
 PF  
 XX 15-NOV-1999; 99US-0164928P.  
 PR  
 XX (BIOM-) BIOMIRA INC.  
 PA  
 XX Jiang Z, Bach M, Yalamati D, Koganty R, Longenecker M;  
 PI WPI; 2001-432554/46.  
 DR  
 XX New synthetic bacterial lipid A analogs, useful e.g. as adjuvants to  
 PT enhance immune responses to antigens in vaccine formulations and as  
 PT anticancer agents.  
 XX  
 PS Disclosure; Fig 35; 155pp; English.  
 XX  
 CC This sequence represents a lipid A analogue of the invention. The

CC analogues of the invention, can also be mucin (MUC1) analogues. The  
 CC analogues can be used as mono- and disaccharide based mimics of bacterial  
 CC lipid A having e.g. one phosphate group at the 4-position as opposed to  
 CC natural lipid A having two phosphate groups at 1- and 4-positions.  
 CC Bacterial lipid A compositions are used as adjuvants to enhance the  
 CC immune responses to various antigens used in vaccine formulations. They  
 CC may also be used as anti-tumour agents, LPS/Lipid A antagonists,  
 CC inhibitors of lipid-A biosynthesis and as antibiotics. They can also be  
 CC used for producing liposomal formulations for treating cancer where the  
 CC liposomal membrane contains the analogues and at least one B-cell or T-  
 CC cell epitope. The synthetic bacterial lipid A analogues have much lower  
 CC toxicity than natural lipid A but with adjuvant properties comparable to  
 CC those of natural lipid A. The analogues are chemically defined with a  
 CC single structure which facilitates their tracking and control from  
 CC manufacturing to final formulation. Production of the analogues is cost  
 CC effective and is easily adaptable for commercial scale up while  
 CC maintaining consistency in both quality and performance. Further, ester  
 CC bonds linking fatty acids to the sugar moiety in natural lipid A and  
 CC which are vulnerable to hydrolysis under physiological conditions leading  
 CC to loss of lipid chains with consequent loss of activity as an adjuvant  
 CC and reduction in shelf life of vaccine formulations, are replaced by  
 CC stable ether (optionally in combination with stable ester) linkages in  
 CC the analogs, which enhances stability and results in longer shelf life  
 XX  
 SQ Sequence 19 AA;

Query Match 100.0%; Score 83; DB 4; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSAPDTRPAGSTA 16  
 Db 1 GVTSAPDTRPAGSTA 16

RESULT 14

ID AAB83310 standard; peptide; 19 AA.

XX AAB83310;

DT 18-SEP-2001 (first entry)

DE Lipopeptide BPI-223.

XX Lipopeptide; Lipid A; MUC1; mucin; analogue; immune response; vaccine;  
 KW anti-tumour agent; antibiotic; cancer; therapy.

XX Synthetic.

Key	Location/Qualifiers
FT Modified-site 3	/note= "T(Tn)"
FT Modified-site 4	/note= "S(Tn)"
FT Modified-site 14	/note= "S(Tn)"
FT Modified-site 15	/note= "T(Tn)"
FT Modified-site 17	/note= "S(Lipo)"
FT Modified-site 18	/note= "S(Lipo)"

XX WO200136433-A2.

PD 25-MAY-2001.

XX 15-NOV-2000; 2000WO-US031281.

XX 15-NOV-1999; 99US-0164928P.

PA (BIOM-) BIOMIRA INC.

XX Jiang Z, Bach M, Yalamati D, Koganty R, Longenecker M;  
 PI WPI; 2001-432554/46.  
 XX  
 PT New synthetic bacterial lipid A analogs, useful e.g. as adjuvants to  
 PT enhance immune responses to antigens in vaccine formulations and as  
 PT anticancer agents.  
 XX  
 PS Disclosure; Fig 35; 155pp; English.  
 XX  
 CC This sequence represents a lipid A analogue of the invention. The  
 CC analogues of the invention, can also be mucin (MUC1) analogues. The  
 CC analogues can be used as mono- and disaccharide based mimics of bacterial  
 CC lipid A having e.g. one phosphate group at the 4-position as opposed to  
 CC natural lipid A having two phosphate groups at 1- and 4-positions.  
 CC Bacterial lipid A compositions are used as adjuvants to enhance the  
 CC immune responses to various antigens used in vaccine formulations. They  
 CC may also be used as anti-tumour agents, LPS/Lipid A antagonists,  
 CC inhibitors of lipid-A biosynthesis and as antibiotics. They can also be  
 CC used for producing liposomal formulations for treating cancer where the  
 CC liposomal membrane contains the analogues and at least one B-cell or T-  
 CC cell epitope. The synthetic bacterial lipid A analogues have much lower  
 CC toxicity than natural lipid A but with adjuvant properties comparable to  
 CC those of natural lipid A. The analogues are chemically defined with a  
 CC single structure which facilitates their tracking and control from  
 CC manufacturing to final formulation. Production of the analogues is cost  
 CC effective and is easily adaptable for commercial scale up while  
 CC maintaining consistency in both quality and performance. Further, ester  
 CC bonds linking fatty acids to the sugar moiety in natural lipid A and  
 CC which are vulnerable to hydrolysis under physiological conditions leading  
 CC to loss of lipid chains with consequent loss of activity as an adjuvant  
 CC and reduction in shelf life of vaccine formulations, are replaced by  
 CC stable ether (optionally in combination with stable ester) linkages in  
 CC the analogs, which enhances stability and results in longer shelf life  
 XX  
 SQ Sequence 19 AA;

Query Match 100.0%; Score 83; DB 4; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSAPDTRPAGSTA 16  
 Db 1 GVTSAPDTRPAGSTA 16

RESULT 15

AAB83308

ID AAB83308 standard; peptide; 19 AA.

XX AAB83308;

DT 18-SEP-2001 (first entry)

DE Lipopeptide BPI-219.

XX Lipopeptide; Lipid A; MUC1; mucin; analogue; immune response; vaccine;  
 KW anti-tumour agent; antibiotic; cancer; therapy.

XX Synthetic.

Key	Location/Qualifiers
FT Modified-site 8	/note= "T(Tn)"
FT Modified-site 17	/note= "S(Lipo)"
FT Modified-site 18	/note= "S(Lipo)"

XX WO200136433-A2.

XX 25-MAY-2001.



XX 15-NOV-2000; 2000WO-US031281.  
 XX  
 XX 15-NOV-1999; 99US-0164928P.  
 XX (BIOM-) BIOMIRA INC.  
 XX  
 XX Jiang Z, Bach M, Yalamati D, Koganty R, Longenecker M;  
 XX WPI; 2001-432554/46.  
 XX  
 XX New synthetic bacterial lipid A analogs, useful e.g. as adjuvants to  
 PT enhance immune responses to antigens in vaccine formulations and as  
 PT anticancer agents.  
 XX  
 XX Disclosure; Fig 35; 155pp; English.  
 XX  
 XX This sequence represents a lipid A analogue of the invention. The  
 CC analogues of the invention, can also be mucin (MUC1) analogues. The  
 CC analogues can be used as mono- and disaccharide based mimics of bacterial  
 CC lipid A having e.g. one phosphate group at the 4-position as opposed to  
 CC natural lipid A having two phosphate groups at 1- and 4-positions.  
 CC Bacterial lipid A compositions are used as adjuvants to enhance the  
 CC immune responses to various antigens used in vaccine formulations. They  
 CC may also be used as anti-tumour agents, LPS/Lipid A antagonists,  
 CC inhibitors of Lipid-A biosynthesis and as antibiotics. They can also be  
 CC used for producing liposomal formulations for treating cancer where the  
 CC liposomal membrane contains the analogues and at least one B-cell or T-  
 CC cell epitope. The synthetic bacterial lipid A analogues have much lower  
 CC toxicity than natural lipid A but with adjuvant properties comparable to  
 CC those of natural lipid A. The analogues are chemically defined with a  
 CC single structure which facilitates their tracking and control from  
 CC manufacturing to final formulation. Production of the analogues is cost  
 CC effective and is easily adaptable for commercial scale up while  
 CC maintaining consistency in both quality and performance. Further, ester  
 CC bonds linking fatty acids to the sugar moiety in natural lipid A and  
 CC which are vulnerable to hydrolysis under physiological conditions leading  
 CC to loss of lipid chains with consequent loss of activity as an adjuvant  
 CC and reduction in shelf life of vaccine formulations, are replaced by  
 CC stable ether (optionally in combination with stable ester) linkages in  
 CC the analogs, which enhances stability and results in longer shelf life  
 XX  
 XX Sequence 19 AA;  
 SQ  
 Query Match 100.0%; Score 83; DB 4; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTAPDTRPAPGSTA 16  
 |||||  
 Db 1 GVTAPDTRPAPGSTA 16  
 |||||  
 RESULT 16  
 AAU07427  
 ID AAU07427 standard; peptide; 19 AA.  
 XX  
 XX AAU07427;  
 AC  
 XX 18-DEC-2001 (first entry)  
 DT  
 XX Mucin MUC-1-derived peptide, BPI-217.  
 DE  
 XX MUC-1; mucin; cytostatic; virucide; vaccine; adjuvant peptide; antigen;  
 KW immune response; cancer; viral disease.  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FH Modified-site 17 /note= "Ser-myristyl"  
 FT Modified-site 18 /note= "Ser-myristyl"  
 FT

XX WO200170265-A2.  
 EN  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-IB000703.  
 XX  
 XX 24-MAR-2000; 2000US-0191736P.  
 PR  
 XX (BIOM-) BIOMIRA INC.  
 XX  
 XX Agrawal B, Longenecker MB, Parker J;  
 PI WPI; 2001-611447/70.  
 XX  
 XX Novel vaccine for treating and preventing disorders such as cancer and  
 PT viral diseases, comprises mucin MUC-1-based adjuvant peptide and an  
 PT antigen.  
 XX  
 XX Claim 5; Page 6; 22pp; English.  
 PS  
 XX The invention relates to a vaccine composition comprising a mucin MUC-1-  
 CC based adjuvant peptide and an antigen. The vaccine composition is useful  
 CC for stimulating the immune response of a patient, by administering the  
 CC vaccine. Alternatively, this can be done by contacting ex vivo a T-cell  
 CC from the patient with the vaccine and administering the contacted cells  
 CC to the patient which stimulates the immune system of the patient. The  
 CC vaccine is useful for treating and preventing disorders such as cancer  
 CC and viral diseases. The vaccine is effective in generating an immune  
 CC response to an antigen against which the patient does not respond. The  
 CC present sequence represents the amino acid sequence of mucin MUC-1-  
 CC derived peptide, BPI-217, which was used in the vaccine composition  
 CC  
 XX Sequence 19 AA;  
 SQ  
 Query Match 100.0%; Score 83; DB 4; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTAPDTRPAPGSTA 16  
 |||||  
 Db 1 GVTAPDTRPAPGSTA 16  
 |||||  
 RESULT 17  
 AAU07431  
 ID AAU07431 standard; peptide; 19 AA.  
 XX  
 XX AAU07431;  
 AC  
 XX 18-DEC-2001 (first entry)  
 DT  
 XX Mucin MUC-1-derived peptide, BPI-216.  
 DE  
 XX MUC-1; mucin; cytostatic; virucide; vaccine; adjuvant peptide; antigen;  
 KW immune response; cancer; viral disease.  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FH Modified-site 3 /note= "O-linked Tn carbohydrate antigen"  
 FT Modified-site 4 /note= "O-linked Tn carbohydrate antigen"  
 FT Modified-site 17 /note= "Ser-myristyl"  
 FT Modified-site 18 /note= "Ser-myristyl"  
 FT  
 XX WO200170265-A2.  
 PN  
 XX 27-SEP-2001.  
 PD  
 XX



PF 23-MAR-2001; 2001WO-IB0000703.  
 XX  
 PR 24-MAR-2000; 2000US-0191736P.  
 XX  
 PA (BIOM-) BIOMIRA INC.  
 XX  
 PI Agrawal B, Longenecker MB, Parker J;  
 XX  
 DR WPI; 2001-611447/70.  
 XX  
 PT Novel vaccine for treating and preventing disorders such as cancer and  
 PT viral diseases, comprises mucin MUC-1-based adjuvant peptide and an  
 PT antigen.  
 XX  
 PS Claim 9; Page 6; 22pp; English.  
 XX  
 CC The invention relates to a vaccine composition comprising a mucin MUC-1-  
 CC based adjuvant peptide and an antigen. The vaccine composition is useful  
 CC for stimulating the immune response of a patient, by administering the  
 CC vaccine. Alternatively, this can be done by contacting ex vivo a T-cell  
 CC from the patient with the vaccine and administering the contacted cells  
 CC to the patient which stimulates the immune system of the patient. The  
 CC vaccine is useful for treating and preventing disorders such as cancer  
 CC and viral diseases. The vaccine is effective in generating an immune  
 CC response to an antigen against which the patient does not respond. The  
 CC present sequence represents the amino acid sequence of mucin MUC-1-  
 CC derived peptide, BPI-216, which was used in the vaccine composition  
 XX  
 SQ Sequence 19 AA;  
 Query Match 100.0%; Score 83; DB 4; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 |||||  
 DB 1 GVTSAPDTRPAPGSTA 16  
 |||||  
 RESULT 18  
 AAE33944  
 ID AAE33944 standard; peptide; 19 AA.  
 XX  
 AC AAE33944;  
 XX  
 DT 02-MAY-2003 (first entry)  
 XX  
 DE MUC 1 peptide, BPI-217.  
 XX  
 KW Liposome; vaccine; immune response; MUC 1 lipopeptide; immunomodulator;  
 KW T cell proliferation; antibody production.  
 XX  
 OS Unidentified.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 17 /note= "Linked to myristyl group"  
 FT Modified-site 18 /note= "Linked to myristyl group"  
 FT  
 XX  
 PN WO200276485-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB002188.  
 XX  
 PR 27-MAR-2001; 2001US-0278698P.  
 XX  
 PA (BUDZ/) BUDZYNSKI W A.  
 XX  
 PI Budzynski WA, Koganty RR, Krantz MJ, Longenecker MB;  
 PI WPI; 2003-046750/04.  
 DR

XX New liposome vaccines comprising at least one monolipopeptide and at  
 PT least one dilipopeptide, useful for modulating the immune response in  
 PT vivo, particularly humoral and cellular immune responses.  
 XX  
 PS Example 2; Page 26; 51pp; English.  
 XX  
 CC The present invention relates to liposomal compositions comprising at  
 CC least one liposome that comprises at least one monolipopeptide and at  
 CC least one dilipopeptide derived from a protein associated with a disease  
 CC selected from the group consisting of tuberculosis, malaria, cancer and  
 CC hepatitis B. The monolipopeptide or dilipopeptide is designed from MUC 1  
 CC protein. The composition is useful as a vaccine for modulating the immune  
 CC response to the peptide in vivo, particularly humoral and cellular immune  
 CC responses where the relative amounts of monolipopeptide and dilipopeptide  
 CC modulate the relative intensities of T cell proliferation and antibody  
 CC production. The present sequence is MUC 1 peptide  
 XX  
 SQ Sequence 19 AA;  
 Query Match 100.0%; Score 83; DB 6; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 |||||  
 DB 1 GVTSAPDTRPAPGSTA 16  
 |||||  
 RESULT 19  
 AAR77000  
 ID AAR77000 standard; peptide; 20 AA.  
 XX  
 AC AAR77000;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 28-FEB-1996 (first entry)  
 XX  
 DE Human mucin 1 repeat unit (MUC1 VNTR)-derived synthetic peptide.  
 XX  
 KW Mucin; repeat unit; VNTR; conjugate; immunogenic; vaccine; allergy;  
 KW therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN EP659768-A2.  
 XX  
 PD 28-JUN-1995.  
 XX  
 PF 26-MAY-1994; 94EP-00303817.  
 XX  
 PR 24-DEC-1993; 93AU-00003223.  
 XX  
 PA (AUST-) AUSTIN RES INST.  
 PA (ILEX-) ILEXUS PTY LTD.  
 XX  
 PI Mckenzie IPC, Apostolopoulos V, Pietersz GA;  
 XX  
 DR WPI; 1995-226205/30.  
 XX  
 PT New conjugates of antigens and carbohydrate polymers - used as  
 PT immunogenic vaccines against disease states, e.g. tumours, HIV infection  
 PT or hepatitis.  
 XX  
 PS Example 1; Page 8; 34pp; English.  
 XX  
 CC AAR77000 is a human mucin repeat unit 1 (MUC1 VNTR) derived peptide. It  
 CC is used as an antigen in a new method for producing an immunogenic  
 CC vaccine. Mucin fragments are cloned into an expression vector and  
 CC expressed so as to produce a glutathione-S-transferase (GST)/mucin  
 CC (MUC) fusion protein. The fusion protein is then conjugated to a  
 CC carbohydrate polymer pref. mannan and it is this cpd.that is used in the  
 CC vaccine. The antigen used does not have to be a mucin peptide it may be

CC obtained from pollen, HIV proteins, hepatitis proteins, tumour virus  
CC proteins or many other viral proteins. Similarly the carbohydrate polymer  
CC may be a polymer from a variety of different sources, e.g. glucose,  
CC galactose, mannose, xylose, glucosamine, fucose, fructose, iduronate 6-O-  
CC methyl-D-galactose, alpha-D-galactopyranose-6-sulphate, mannuronate or  
CC rhamnose. The cpds. provoke a potentiated cellular response of activated  
CC T-lymphocytes which are cytotoxic to cells expressing the antigen  
CC component and as such are useful for producing vaccines against many  
CC disease states such as viral tumours. Other diseases against which the  
CC cpds. may act as immunogenic vaccines are: type I allergies; malaria; HIV  
CC ; dental caries; foot and mouth disease; meningitis; whooping cough;  
CC rabies; tuberculosis; hepatitis; herpes and the common cold. (Updated on  
CC 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 20 AA;

Query Match 100.0%; Score 83; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||||  
Db 4 GVTSAPDTRPAPGSTA 19

RESULT 20  
AAW21976  
ID AAW21976 standard; peptide; 20 AA.

XX AAW21976;

XX 30-OCT-1997 (first entry)

XX Human MUC1 VNTR peptide Muc pep2.

DE Cancer; vaccine; peptide mimic; mucin; MUC1; Gal alpha(1,3)Gal;  
KW immunotherapy; therapy; adenocarcinoma; VNTR;  
KW variable number of tandem repeat.

OS Homo sapiens.

XX WO9711715-A1.

PN 03-APR-1997.

XX 27-SEP-1996; 96WO-AU000617.

XX 27-SEP-1995; 95AU-00005680.

XX (AUST-) AUSTIN RES INST.

XX Sandrin MS, McKenzie IFC, Apostolopoulos V;

XX WPI; 1997-212670/19.

XX Cancer vaccine containing MUC1 peptide mimic - used in the treatment of  
PT adenocarcinoma.

PS Example 1; Page 27; 55pp; English.

XX Muc pep2 is a peptide derived from the variable number of tandem repeat  
CC (VNTR) region (residues 13-32) of human cancer mucin MUC1. It is capable  
CC of binding to anti-Gal alpha(1,3)Gal antibody. A novel cancer vaccine  
CC comprises a peptide (AAW21680-86) which mimics MUC1 or other cancer  
CC peptides and one or more pharmaceutically acceptable carrier or diluent,  
CC optionally in association with an appropriate carrier peptide or other  
CC therapeutic agent. Mucin VNTR peptides (AAW21975-85) may also be used.  
CC The vaccine is used in the treatment of a patient suffering from, or with  
CC a predisposition to, adenocarcinoma (claimed)

XX Sequence 20 AA;

Query Match 100.0%; Score 83; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||||  
Db 4 GVTSAPDTRPAPGSTA 19

RESULT 21

AAW22250

ID AAW22250 standard; peptide; 20 AA.

XX AAW22250;

XX 03-DEC-1997 (first entry)

XX Muc pep 2, anti-Gal-alpha(1,3)Gal antibody binding peptide.

DE Consensus sequence; antibody binding peptide; xenotransplant rejection;  
XX cell surface protein; anti-Gal-alpha(1,3)Gal antibody; Gal-alpha(1,3)Gal;  
KW endothelial cell; complement cascade; coagulation cascade; immunogen;  
KW hyperacute rejection; human; xenograft organ; therapy; mouse.

OS Homo sapiens.

XX WO9711963-A1.

PN 03-APR-1997.

XX 27-SEP-1996; 96WO-AU000616.

XX 27-SEP-1995; 95AU-00005680.

XX (AUST-) AUSTIN RES INST.

XX Sandrin MS, McKenzie IFC, Gallop MA;

XX WPI; 1997-212848/19.

XX Peptide capable of binding to anti-Gal-alpha(1,3)Gal antibodies - useful  
PT for preventing hyper-acute xeno-transplant rejection in humans.

PS Example 2; Page 26; 49pp; English.

XX AAW22245-W22260 represent examples of peptides of the invention. The  
CC peptides of the invention are capable of binding to anti-Gal-  
CC alpha(1,3)Gal (galactose in an alpha(1,3) linkage with galactose)  
CC antibodies. The peptides of the invention contain a consensus sequence  
CC represented by one of the peptides shown in AAW22238-W22244. The binding  
CC of the anti-Gal-alpha(1,3)Gal antibody to the Gal-alpha(1,3)Gal epitope  
CC on the surface of endothelial cells results in the activation of both the  
CC complement and coagulation cascade, resulting in hyperacute rejection.  
CC Higher primates do not make Gal-alpha(1,3)Gal but humans do possess  
CC natural antibodies to Gal-alpha(1,3)Gal(15,18). The binding peptides when  
CC in association with other immunoassay components are useful for the  
CC detection of anti-Gal-alpha(1,3)Gal antibodies in human biological  
CC samples. The binding peptides are useful for blocking immunogenic  
CC activity of human anti-Gal-alpha(1,3)Gal antibodies. They are also useful  
CC for preventing hyperacute xenotransplant rejection in humans by  
CC administration of the peptide to the patient who is to receive a donor  
CC animal organ, in particular the peptide is administered as a vaccine and  
CC induces immunological tolerance to the Gal-alpha(1,3)Gal epitope. Solid  
CC substrates which have the binding peptide immobilised to them, are useful  
CC for the removal of anti-Gal(1,3) antibodies from plasma of human patients  
CC who will receive a xenograft organ, by passing the plasma over the  
CC substrate and then returning the plasma to the patient

XX Sequence 20 AA;

Query Match 100.0%; Score 83; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GVTAPDTRPAGSTA 16
Db 4 GVTAPDTRPAGSTA 19

RESULT 22
AAW27393
ID AAW27393 standard; peptide; 20 AA.
XX AC AAW27393;
XX DT 14-APR-1998 (first entry)
XX DE Human epithelial cell mucin MUC1 modified epitope.
XX KW Hepatitis B virus; HBV; HBcAg; tumour antigen epitope; chimeric;
XX XW Ha-ras oncogene; epithelial cell mucin; MUC1; human; tumour cell growth.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 6.10
XX FT /note= "immunogenic sequence"
XX PN W09735008-A1.
XX PD 25-SEP-1997.
XX PF 21-MAR-1997; 97WO-US004656.
XX PR 21-MAR-1996; 96US-0013839P.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Kwak LW, Biragyn A;
XX DR WPI; 1997-480214/44.
XX CC Chimeric polypeptide comprising hepatitis B virus core antigen and tumour
XX PT antigen epitope - useful to inhibit tumour cell growth.
XX PS Example 1; Page 22; 50pp; English.
XX CC This is a modified epitope of the human epithelial cell mucin MUC1. The
XX CC immunogenic sequence is located internally in this modified MUC1 epitope.
XX CC This is used in the construction of a chimeric polypeptide comprising a
XX CC Hepatitis B virus core antigen (HBcAg) and a tumour antigen epitope. The
XX CC tumour antigen epitope is from the human epithelial cell mucin or is from
XX CC the Ha-ras oncogene. The polypeptide can be used to inhibit the growth of
XX CC tumour cells, especially by inducing an immune response to generate
XX CC antibodies against a pathological or harmful condition
XX SQ Sequence 20 AA;
XX Query Match 100.0%; Score 83; DB 2; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 5.7e-05;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAGSTA 16
Db 1 GVTAPDTRPAGSTA 16

RESULT 23
AAW25112
ID AAY25112 standard; peptide; 20 AA.
XX AC AAY25112;
XX DT 25-AUG-1999 (first entry)
XX DE Synthetic glycopeptide #2 homologous to MUC1.
XX
```

```
KW Immunodominant; epitope; MUC1; epithelial mucin; tumour; vaccine;
KW mammary carcinoma; colorectal carcinoma; pancreas carcinoma.
OS Synthetic.
XX DE19758400-A1.
XX 01-JUL-1999.
XX 30-DEC-1997; 97DE-01058400.
XX 30-DEC-1997; 97DE-01058400.
XX {DELB-} DELBRUECK CENT MOLEKULARE MEDIZIN MAX.
XX {HANI-} HANISCH F.
XX Karsten U, Hanisch F, Paulsen H;
XX WPI; 1999-372374/32.
XX New tumour vaccine - for treating tumour cells of mammary tumours,
XX colorectal and pancreas carcinomas.
XX Example 1; Page 3; 6pp; German.
XX CC This invention describes a novel tumour vaccine which comprises a
XX CC synthetic peptide derived from human epithelial mucin MUC1. This active
XX CC tumour vaccine can be used against tumour cells from mammary, colorectal
XX CC or pancreas carcinomas. This sequence represents a synthetic glycopeptide
XX CC which is homologous to an immunodominant epitope from human MUC1
XX SQ Sequence 20 AA;
XX Query Match 100.0%; Score 83; DB 2; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 5.7e-05;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAGSTA 16
Db 2 GVTAPDTRPAGSTA 17

RESULT 24
AAW67589
ID AAW67589 standard; peptide; 20 AA.
XX AC AAW67589;
XX DT 02-MAR-1999 (first entry)
XX DE T-cell activation peptide #7.
XX KW Activated T helper cell; CD4+; cytotoxic T cell; CD8+; liposome; epitope;
XX KW peripheral blood lymphocyte; antigen-presenting cell; APC; virus; tumour;
XX KW bacterium; parasite; cytokine; vaccine; cancer; malaria; HIV; hepatitis;
XX KW tuberculosis.
XX OS Synthetic.
XX PN W09850527-A1.
XX PD 12-NOV-1998.
XX PF 07-MAY-1998; 98WO-US009288.
XX PR 08-MAY-1997; 97US-0045949P.
XX PA (BIOM-) BIOMIRA INC.
XX PI Agrawal B, Krantz MJ, Reddish MA, Longenecker BM;
XX DR WPI; 1999-034715/03.
XX
```

PT Method of activation of T cells - by exposure to antigen-presenting cells  
 PT loaded with antigen in liposome, used for, e.g. treating cancer and  
 PT microbial infections.

XX Claim 14; Page 43; 75pp; English.

CC Peptides AAW67583-W67611 are used to produce activated T helper (CD4+) and cytotoxic (CD8+) T-cells. The activated T cells are produced by treating peripheral blood lymphocytes with liposome-encapsulated peptide antigen to generate Ag-loaded antigen-presenting cells (APC), contacting naive or anergic T-cells with these APC, and isolating the resulting activated T-cells. The cells are specific for a particular antigen, particularly one derived from a tumour, but also those from viruses, bacteria and other parasites. It can also be used to identify antigens and epitopes able to generate an Ag-specific T-cell response (by assessing proliferation and cytokine release). Also the Ag-loaded APC can be used as cellular vaccines for treating cancer (claimed) or other diseases (e.g. malaria, human immune deficiency virus infection, hepatitis, tuberculosis). The activated T-cells can be used to treat the same conditions by adoptive T-cell transfer therapy

XX Sequence 20 AA;

Query Match 100.0%; Score 83; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
 ID AAY80111 standard; peptide; 20 AA.  
 DB 1 GVTSAPDTRPAPGSTA 16

RESULT 25

AAAY80111  
 ID AAY80111 standard; peptide; 20 AA.

XX AC AAY80111;

DT 17-MAY-2000 (first entry)

DE MUC-1 repeat peptide sequence.

XX Human; MUC-1; detection; T-cell activation; mucin; antiinflammatory;  
 KW immunomodulator; antirheumatic; antiarthritic; antiallergic;  
 KW dermatological; antidiabetic; nephrotropic; antithyroid; antianaemic;  
 KW neuroprotective; hepatotropic; uropathic; ophthalmological; antiviral;  
 KW cytostatic; autoimmune disorder; inflammatory disorder; viral disease;  
 KW cancer.

XX Homo sapiens.

OS WO200000828-A1.

PN 06-JAN-2000.

ED 25-JUN-1999; 99WO-US012820.

PF 26-JUN-1998; 98US-0090916P.

PR (BIOM-) BIOMIRA INC.

XX Agrawal B, Longenecker BM;

PI WPI; 2003-170935/15.

XX Detecting T-cell activation by measuring the amount of MUC-1 expression useful for diagnosing or treating autoimmune or inflammatory disorders, viral disease or cancer.

XX Disclosure; Page 11; 40pp; English.

XX A method has been developed for detecting T-cell activation by evaluating the amount of MUC-1 mucin expression in a T-cell compared to a non-

CC activated control. The method is useful for treating disorders associated with T-cell activation, using an agent (antibody/antagonist) that modulates MUC-1 activity. The T-cell activation associated disorders may be autoimmune or inflammatory disorders (e.g. inflammatory arthritis, rheumatoid arthritis, psoriasis, allergies, allergic contact dermatitis, ankylosing spondylitis, myasthenia gravis, systemic lupus erythematosus, polyarthritis nodosa, Goodpastures syndrome, isopathic thrombocytopenic purpura, autoimmune haemolytic anaemia, Grave's disease, rheumatic fever, pernicious anaemia, insulin-resistant diabetes mellitus, bullous pemphigus vulgaris, viral myocarditis (Cocksakie B virus response), autoimmune thyroiditis (Hashimoto's disease), male infertility (autoimmune), sarcoidosis, allergic encephalomyelitis, multiple sclerosis, Sjorgens disease, Reiter's disease, Celiac disease, sympathetic ophthalmia, and primary biliary cirrhosis), viral disease or cancer. The present sequence represents a MUC-1 peptide from the present invention

XX Sequence 20 AA;

Query Match 100.0%; Score 83; DB 3; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
 ID AAB09915 standard; peptide; 20 AA.  
 DB 1 GVTSAPDTRPAPGSTA 16

RESULT 26

AAAB09915  
 ID AAB09915 standard; peptide; 20 AA.

XX AC AAB09915;

DT 06-NOV-2000 (first entry)

DE MUC-1 derivative peptide #1.

XX Targeting signal; MUC-1; immunosuppression; autoimmune disorder;  
 KW immune disorder; inflammatory disorder.

XX Synthetic.

PN WO200034468-A2.

XX 15-JUN-2000.

PF 09-DEC-1999; 99WO-US029016.

XX 11-DEC-1998; 98US-0111973P.

XX (BIOM-) BIOMIRA INC.

PI Agrawal B, Longenecker BM;

XX WPI; 2000-423418/36.

XX Use of agent capable of intracellularly inhibiting mucin MUC-1 for inducing T-cell-based immunosuppression and for treating autoimmune disorders, transplant rejection and inflammatory disorders.

XX Disclosure; Page 16; 51pp; English.

XX The present sequence is an antagonistic peptide derived from the MUC-1 protein. It can be joined to a targeting sequence to internalise it within the cell. MUC-1 is an immunosuppressor, and antagonists such as this act to reduce overactive immune responses. Thus, the peptide can be used to treat inflammatory disorders such as rheumatoid arthritis, psoriasis, allergic contact dermatitis and ankylosing spondylitis, autoimmune disorders including myasthenia gravis, systemic lupus erythematosus, polyarthritis nodosa, Goodpastures syndrome, isopathic thrombocytopenic purpura, autoimmune haemolytic anaemia, Graves' disease, rheumatic fever, pernicious anaemia, insulin-resistant diabetes

CC mellitus, bullous pemphigoid, pemphigus vulgaris, viral myocarditis,  
CC autoimmune thyroiditis, male infertility, sarcoidosis, allergic  
CC encephalomyelitis, multiple sclerosis, Sjorgens disease, Reiter's  
CC disease, Celiac disease, sympathetic ophthalmia and primary biliary  
CC cirrhosis, immune disorders, graft versus host disease and transplant  
CC rejection  
XX  
SQ Sequence 20 AA;  
  
Query Match 100.0%; Score 83; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 GVTSAPDTRPAPGSTA 16  
Db 1 GVTSAPDTRPAPGSTA 16  
|||||  
  
RESULT 27  
AAB29928  
ID AAB29928 standard; peptide; 20 AA.  
XX  
AC AAB29928;  
XX  
DT 12-FEB-2001 (first entry)  
DE Cytochemical assay quality control CA15-3 peptide SEQ ID NO: 1.  
XX  
KW Cytochemical assay; quality control; immunohistochemical stain; cancer;  
KW viral infection.  
XX  
OS Synthetic.  
XX  
PN WO200062064-A2.  
XX  
PD 19-OCT-2000.  
XX  
PF 14-APR-2000; 2000WO-US010268.  
XX  
PR 14-APR-1999; 99US-00291351.  
XX  
PA (CYTO-) CYTOLOGIX CORP.  
XX  
PI Bogen SA, Radcliffe GE, Sompuram SR, Ramanathan H;  
XX WPI; 2000-679514/66.  
XX  
PT Quality control device for cytochemical assays measuring analytes in a  
PT sample, comprises matrix with top surface having a quality control  
PT reagent moiety and a bottom adhesive surface for adhering device to  
PT platform.  
XX  
PS Disclosure; Page 21; 75pp; English.  
XX  
CC The present invention is related to a device useful for measuring the  
CC quality of cytochemical assays such as immunohistochemical stains. These  
CC assays can be used in cancer diagnosis, and in the diagnosis of viral  
CC infections. The quality control device comprises a matrix with a top  
CC surface having a quality control reagent moiety and the bottom surface  
CC having an adhesive to enable the adherence of the device to a planar test  
CC platform. The present sequence was used in assays to demonstrate the  
CC device  
XX  
SQ Sequence 20 AA;  
  
Query Match 100.0%; Score 83; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 GVTSAPDTRPAPGSTA 16  
Db 4 GVTSAPDTRPAPGSTA 19  
|||||  
|||||

RESULT 28  
AAE12708  
ID AAE12708 standard; peptide; 20 AA.  
XX  
AC AAE12708;  
XX  
DT 04-JAN-2002 (first entry)  
XX  
DE Human MUC1 core peptide fragment.  
XX  
KW Human; tumour-associated antigen mucin-1; MUC-1; adenocarcinoma; VL;  
KW variable light chain region; cancer; breast; ovary; lung; bladder;  
KW cytostatic; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200175110-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US010589.  
XX  
PR 30-MAR-2000; 2000US-00538913.  
XX  
PA (DYAX-) DYAX CORP.  
XX  
PI Hoogenboom HRJM, Henderikx MPG;  
XX WPI; 2001-626437/72.  
XX  
PT Novel isolated tumor-associated antigen mucin-1-specific binding member  
PT for diagnosing and treating cancer, comprises mucin-1 binding domain or  
PT its portion for binding to an epitope of the protein core of mucin-1.  
XX  
PS Example 1; Page 32; 126pp; English.  
XX  
CC The invention relates to an isolated tumour-associated antigen mucin-1  
CC (MUC-1)-specific binding member comprising an antigen binding domain  
CC region having an antibody variable light (VL) or heavy (VH) region, or a  
CC complementarity determining region (CDR) of VL or VH. MUC1-specific  
CC binding member is useful for diagnosing cancer, preferably adenocarcinoma  
CC The binding of MUC1-specific binding member to MUC1 is detected by a  
CC detection method selected from enzyme-linked immunosorbent assay,  
CC magnetic resonance imaging, scintillation counting, and X-ray film. MUC1-  
CC specific binding member is useful for treating cancer, preferably  
CC adenocarcinoma, in an individual, where the cancer is present in tissue  
CC of the breast, ovary, lung, or bladder of the individual. MUC1-specific  
CC binding member is useful for diagnosing and imaging MUC1-expressing  
CC cancer cells and tissues, for purifying or isolating non-glycosylated,  
CC underglycosylated or cancer-associated forms of MUC1 or MUC1 epitope-  
CC containing molecules, and for therapeutically or prophylactically  
CC treating cancer. The present sequence is human MUC1 core peptide fragment  
XX  
SQ Sequence 20 AA;  
  
Query Match 100.0%; Score 83; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 GVTSAPDTRPAPGSTA 16  
Db 4 GVTSAPDTRPAPGSTA 19  
|||||  
|||||



DE Mucin, MUC1, VNTR monomer.  
 XX Human; Mucin; MUC1; VNTR; nucleic acid vaccine; breast cancer;  
 KW epithelial cell tumour; immunogen.  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 6 /note= "O-glycosylated"  
 FT Modified-site 7 /note= "O-glycosylated"  
 FT Peptide 8..12 /label= Epitope  
 FT Modified-site 11 /note= "O-glycosylated"  
 FT Modified-site 17 /note= "O-glycosylated"  
 FT Modified-site 18 /note= "O-glycosylated"  
 FT  
 XX WO200146228-A2.  
 PN  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 20-DEC-2000; 2000WO-GB004906.  
 XX  
 PR 22-DEC-1999; 99GB-00030359.  
 XX  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI Crowe JS, Ellis JH;  
 XX WPI; 2001-418023/44.  
 DR  
 XX Novel nucleic acid vaccine construct for vaccinating mammals against  
 PT tumors, comprising a polynucleotide which when expressed in mammalian  
 PT cell, expresses a polypeptide with altered glycosylation site.  
 XX  
 PS Example 1; Page 18; 37pp; English.  
 XX  
 CC The sequence represents the mucin, MUC1, VNTR (variable number of tandem  
 CC repeats) monomer. The invention relates to a nucleic acid vaccine  
 CC encoding at least five consecutive amino acids from the VNTR, where one  
 CC or more of the amino acids is a glycosylation site. The nucleic acid  
 CC vaccines expressing the MUC1 VNTRs are useful for vaccination of a mammal  
 CC against tumours e.g., epithelial cell tumours or breast cancer tumours  
 CC and encode polypeptides that retain conformation of MUC1 epitopes, an  
 CC essential requirement for continued immunogenicity of the altered  
 CC polypeptides, and which have reduced glycosylation, hence resembling more  
 CC closely the form of MUC1 expressed on tumours  
 XX  
 SQ Sequence 20 AA;  
 Query Match 100.0%; Score 83; DB 4; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 Db |||||  
 4 GVTSAPDTRPAPGSTA 19  
 RESULT 30  
 AAU82063  
 ID AAU82063 standard; peptide; 20 AA.  
 XX  
 AC AAU82063;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Antigenic peptide M1c associated with cancer (Muc1).  
 XX

KW T-cell binding ligand; TCBL; peptide G'; human MHC class II beta chain;  
 KW peptide J; human beta-2-microglobulin; HIV-1; TCBL peptide construct;  
 KW immunological disorder; immune response; human immunodeficiency virus;  
 KW herpes simplex virus infection; HSV; malaria; tuberculosis; cancer; CEA;  
 KW acquired immunodeficiency syndrome; AIDS; allergy; autoimmune disease;  
 KW autoimmune myocarditis; cytostatic; antiinflammatory.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200189286-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 24-MAY-2001; 2001WO-US016793.  
 XX  
 PR 24-MAY-2000; 2000US-0206548P.  
 XX  
 PA (CELS-) CEL-SCI CORP.  
 XX  
 PI Zimmerman DS, Sarin PS;  
 XX  
 DR WPI; 2002-083037/11.  
 XX  
 PT New T cell binding ligand peptide for treating immunological disorders  
 PT such as herpes simplex virus, tuberculosis, cancers, acquired  
 PT immunodeficiency syndrome and allergies.  
 XX  
 PS Disclosure; Page 26; 110pp; English.  
 XX  
 CC The present invention relates to novel T-cell binding ligand (TCBL)  
 CC peptides (e.g. peptide G' (modified human MHC class II beta chain peptide  
 CC G, peptide J (human beta-2-microglobulin peptide) and HIV-1 peptides) and  
 CC TCBL peptide constructs for treating immunological disorders. The peptide  
 CC constructs are useful for eliciting a cellular immune response in a human  
 CC patient. The method comprises administering the peptide construct to the  
 CC patient preferably in combination with an immune response adjuvant. The  
 CC peptide constructs in the form of conjugated peptides are useful for  
 CC eliciting a cellular immune response in a patient exposed to or at risk  
 CC for exposure to the human immunodeficiency virus (HIV). The TCBL peptides  
 CC are useful for treating a patient suffering from an immunological  
 CC disorder such as herpes simplex virus (HSV) infection, malaria,  
 CC tuberculosis, cancers, acquired immunodeficiency syndrome (AIDS),  
 CC allergies, autoimmune diseases (e.g. arthritis, Graves disease, multiple  
 CC sclerosis (MS), autoimmune myocarditis, diabetes and lupus) by  
 CC administering a peptide construct comprising a TCBL peptide bonded to an  
 CC antigenic peptide associated with the disorder. Unlike prior art peptide  
 CC conjugates, a modified version of peptide G has long range stabilisation  
 CC and also enhances the immune response. AAU82019-AAU82114 represent T-cell  
 CC specific binding ligand peptides, peptide constructs or peptides used in  
 CC their construction  
 XX  
 SQ Sequence 20 AA;  
 Query Match 100.0%; Score 83; DB 5; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 Db |||||  
 1 GVTSAPDTRPAPGSTA 16  
 RESULT 31  
 ABB08278  
 ID ABB08278 standard; peptide; 20 AA.  
 XX  
 AC ABB08278;  
 XX  
 DT 15-AUG-2002 (first entry)  
 XX  
 DE Human cancer-associated mucin MUC1 tandem repeat O-glycosylated region.  
 XX MUC1; human; mucin; cancer; N-acetylgalactosaminyltransferase; tumour;  
 KW



KW GalNac-transferase; cytostatic; antiasthmatic; anti-inflammatory;  
KW immunosuppressive; antiarthritic; vulnery; antibacterial; lung disease;  
XX exocrine gland; leukocyte trafficking.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 4 /note= "O-glycosylated with GalNac"

FT Modified-site 15 /note= "O-glycosylated with GalNac"

FT Modified-site 16 /note= "O-glycosylated with GalNac"

XX WO200185215-A2.

XX 15-NOV-2001.

XX 10-MAY-2001; 2001WO-DK000328.

XX 11-MAY-2000; 2000US-0203331P.

XX (CLAU/) CLAUSEN H.

XX Clausen H, Hassan H, Reis CA, Bennett EP;

XX WPI; 2002-240406/29.

XX Modulating functions of polypeptide N-acetylgalactosaminyltransferases  
PT (GalNac-transferases), useful for treating tumors, cancers, or lung  
PT diseases, by administering a modulator of the lectin domain of GalNac-  
PT transferases.

XX Disclosure; Page 4; 55pp; English.

XX The sequence represents the human cancer-associated mucin MUC1 tandem  
CC repeat O-glycosylated region. The invention relates to a novel method for  
CC modulating functions of polypeptide N-acetylgalactosaminyltransferases  
CC (GalNac-transferases). The method of the invention has cytostatic,  
CC antiasthmatic, anti-inflammatory, immunosuppressive, antiarthritic,  
CC vulnery, and antibacterial activity. The method is useful for  
CC modulating functions of polypeptide GalNac-transferases and for  
CC identifying agents which can modulate or inhibit the functions of  
CC polypeptide GalNac-transferases. Agents which can inhibit one or more  
CC lectin domains of polypeptide GalNac-transferases and modulate functions  
CC mediated by the lectin domains are useful in the treatment of tumors,  
CC cancers, lung diseases, diseases of exocrine glands associated with  
CC increased or decreased mucin secretion, and disorders associated with  
CC deregulation of selectin-mediated leukocyte trafficking. Inhibitors of  
CC lectin domains of GalNac-transferases may be used to manipulate disease-  
CC associated O-glycosylation to augment immunity and to prepare vaccines,  
CC to manipulate mucin secretion and O-glycan density in diseases associated  
CC with mucous accumulation to decrease secretion and enhance clearance of  
CC mucins, and to modulate O-glycosylation of recombinant glycoproteins by  
CC inhibition of polypeptide GalNac-transferases in host expression cells

XX Sequence 20 AA;

XX Query Match 100.0%; Score 83; DB 5; Length 20;

XX Best Local Similarity 100.0%; Pred. No. 5.7e-05;

XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16

Db 2 GVTSAPDTRPAPGSTA 17

XX RESULT 32

XX ABU58800

XX ID ABU58800 standard; peptide; 20 AA.

XX AC ABU58800;

XX XX

DT 15-APR-2003 (First entry)

XX Mucin 1 (MUC-1) peptide #1.

XX Mucin-1-specific binding member; human; cancer; adenocarcinoma;  
KW breast cancer; ovarian cancer; bladder cancer; lung cancer;  
KW anti-cancer regimen; anti-cancer drug; radiation treatment.

XX Homo sapiens.

XX US2002146750-A1.

XX 10-OCT-2002.

XX 30-MAR-2001; 2001US-00822698.

XX 30-MAR-2000; 2000US-00538913.

XX (HOOG/) HOOGENBOOM H R J M.

XX (HEND/) HENDERIKX M P G.

XX Hooogenboom HRJM, Henderikx MPG;

XX WPI; 2003-198057/19.

XX Isolated mucin-1-specific binding member for diagnosing and/or treating  
PT cancer, e.g. breast cancer, comprises antigen binding domain having  
PT region that contains specific amino acid sequence.

XX Disclosure; Page 36; 70pp; English.

XX The invention describes an isolated mucin-1-specific binding member  
CC having an antigen binding domain including a region that comprises a  
CC specific amino acid sequence. The inventive MUC1-specific binding member  
CC is used in the diagnosis and/or treatment of cancer, e.g. adenocarcinoma,  
CC found in various tissues, e.g. breast, ovary, bladder, and lung. It can  
CC be used alone or as a component in a more complex anti-cancer regimen  
CC which may contain anti-cancer drug(s) and/or radiation treatment(s). The  
CC inventive binding member recognizes tumour-associated MUC1 on  
CC adenocarcinoma. Its affinity is high enough to bind to tumour cells. This  
CC is the amino acid sequence of a mucin 1 (MUC-1) peptide used in the  
CC isolation MUC-1 antigen binding domains for use in the treatment of  
CC cancer

XX Sequence 20 AA;

XX Query Match 100.0%; Score 83; DB 6; Length 20;

XX Best Local Similarity 100.0%; Pred. No. 5.7e-05;

XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16

Db 4 GVTSAPDTRPAPGSTA 19

XX RESULT 33

XX AAE33943

XX ID AAE33943 standard; peptide; 20 AA.

XX AC AAE33943;

XX 02-MAY-2003 (first entry)

XX MUC 1 peptide #5.

XX Liposome; vaccine; immune response; MUC 1 lipopeptide; immunomodulator;

XX T cell proliferation; antibody production.

XX Unidentified.

XX WO200276485-A2.

XX 03-OCT-2002.

XX PF 27-MAR-2002; 2002WO-IB002188.  
 XX PR 27-MAR-2001; 2001US-0278698P.  
 XX PA (BUDZ/) BUDZYNSKI W A.  
 XX PI Budzynski WA, Koganty RR, Krantz MJ, Longenecker MB;  
 XX DR WPI; 2003-046750/04.  
 XX New liposome vaccines comprising at least one monolipopeptide and at  
 PT least one dilipopeptide, useful for modulating the immune response in  
 PT vivo, particularly humoral and cellular immune responses.  
 XX PS Disclosure; Page 21; 51pp; English.  
 XX CC The present invention relates to liposomal compositions comprising at  
 CC least one liposome that comprises at least one monolipopeptide and at  
 CC least one dilipopeptide derived from a protein associated with a disease  
 CC selected from the group consisting of tuberculosis, malaria, cancer and  
 CC hepatitis B. The monolipopeptide or dilipopeptide is designed from MUC 1  
 CC protein. The composition is useful as a vaccine for modulating the immune  
 CC response to the peptide in vivo, particularly humoral and cellular immune  
 CC responses where the relative amounts of monolipopeptide and dilipopeptide  
 CC modulate the relative intensities of T cell proliferation and antibody  
 CC production. The present sequence is MUC 1 peptide  
 XX SQ Sequence 20 AA;  
 Query Match 100.0%; Score 83; DB 6; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 Db 1 GVTSAPDTRPAPGSTA 16  
 RESULT 34  
 AAE33957  
 ID AAE33957 standard; peptide; 20 AA.  
 AC AAE33957;  
 XX 02-MAY-2003 (first entry)  
 XX Dilipidated MUC 1 peptide, BP1-275.  
 DE Liposome; vaccine; immune response; MUC 1 lipopeptide; immunomodulator;  
 KW T cell proliferation; antibody production.  
 XX Unidentified.  
 XX Key Location/Qualifiers  
 FH Modified-site 18 /note= "Linked to lipo group"  
 FT Modified-site 19 /note= "Linked to lipo group"  
 XX WO200276485-A2.  
 XX 03-OCT-2002.  
 XX 27-MAR-2002; 2002WO-IB002188.  
 XX PR 27-MAR-2001; 2001US-0278698P.  
 XX PA (BUDZ/) BUDZYNSKI W A.  
 XX PI Budzynski WA, Koganty RR, Krantz MJ, Longenecker MB;  
 XX DR WPI; 2003-046750/04.

XX New liposome vaccines comprising at least one monolipopeptide and at  
 PT least one dilipopeptide, useful for modulating the immune response in  
 PT vivo, particularly humoral and cellular immune responses.  
 XX Example 7; Page 37; 51pp; English.  
 XX The present invention relates to liposomal compositions comprising at  
 CC least one liposome that comprises at least one monolipopeptide and at  
 CC least one dilipopeptide derived from a protein associated with a disease  
 CC selected from the group consisting of tuberculosis, malaria, cancer and  
 CC hepatitis B. The monolipopeptide or dilipopeptide is designed from MUC 1  
 CC protein. The composition is useful as a vaccine for modulating the immune  
 CC response to the peptide in vivo, particularly humoral and cellular immune  
 CC responses where the relative amounts of monolipopeptide and dilipopeptide  
 CC modulate the relative intensities of T cell proliferation and antibody  
 CC production. The present sequence is dilipidated MUC 1 peptide  
 XX SQ Sequence 20 AA;  
 Query Match 100.0%; Score 83; DB 6; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 Db 1 GVTSAPDTRPAPGSTA 16  
 RESULT 35  
 AAE30200  
 ID AAE30200 standard; peptide; 20 AA.  
 XX AC AAE30200;  
 XX 24-FEB-2003 (first entry)  
 XX Human MUC-1 peptide.  
 DE Human; mucin peptide; MUC-1; vaccine; immune response; tumour; cancer;  
 KW hyperplasia; metaplasia; dysplasia; therapy.  
 XX Homo sapiens.  
 OS WO200262319-A2.  
 XX PD 15-AUG-2002.  
 XX 22-JAN-2002; 2002WO-US001852.  
 PF 19-JAN-2001; 2001US-0262699P.  
 XX (CORI-) CORIXA CORP.  
 XX Johnson ME, Cecil T, Finn OJ;  
 PI WPI; 2003-029829/02.  
 DR Composition, useful for formulating vaccines for treating or preventing  
 XX cancer or a pre-cancerous condition, comprises mucin peptide and  
 PT biodegradable polymeric microsphere.  
 XX Claim 3; Page 30; 42pp; English.  
 XX The invention relates to a composition comprising a mucin peptide such as  
 CC MUC-1 and a biodegradable polymeric microsphere. Composition of the  
 CC invention is used for the production of vaccines which are used for  
 CC stimulating an immune response to MUC-1 in a subject; inhibiting tumour  
 CC growth in a subject having a cancer associated with reduced glycosylation  
 CC of MUC-1; prolonging survival in a subject having cancer associated with  
 CC reduced glycosylation of MUC-1; and treating or preventing cancer  
 CC associated with reduced glycosylation of MUC-1. Vaccines are also useful  
 CC for treating or preventing pre-cancerous condition e.g. hyperplasia,

CC metaplasia or dysplasia. The present sequence is human MUC-1 peptide  
XX  
SQ Sequence 20 AA;  
  
Query Match 100.0%; Score 83; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GVTAPDTRPAPGSTA 16  
Db 1 GVTAPDTRPAPGSTA 16  
  
RESULT 36  
ADB84183  
ID ADB84183 standard; peptide; 20 AA.  
XX  
AC ADB84183;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human MUC1 repeat sequence SEQ ID NO:10.  
XX  
KW human; immune system; immunostimulatory; CpG dinucleotide;  
KW immunostimulant; MUC1.  
XX  
OS Homo sapiens.  
XX  
PN WC2003066649-A1.  
XX  
PD 14-AUG-2003.  
XX  
PF 04-FEB-2003; 2003WO-CA000135.  
XX  
PR 04-FEB-2002; 2002US-0353195P.  
XX  
PA (BIOM-) BIOMIRA INC.  
XX  
PI Jiang Z, Koganty RR, Yalamati D, Baek M;  
XX WPI; 2003-767230/72.  
XX  
PT Stimulation of immune system involves administration of immunostimulatory molecule comprising at least one oligonucleotide strand having at least one nucleotide sequence and at least one covalently incorporated lipophilic group.  
XX  
PS Disclosure; Page 88; 95pp; English.  
XX  
CC The invention relates to a novel method for stimulating an immune system.  
CC The method involves administering an immunostimulatory molecule of covalently lipidated oligonucleotides comprising the CpG dinucleotide unit, or an analogue. The method of the invention has immunostimulant activity. The invention is useful in the manufacture of a composition for immunostimulating a subject. The present sequence is used in the exemplification of the invention.  
XX  
SQ Sequence 20 AA;  
  
Query Match 100.0%; Score 83; DB 7; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GVTAPDTRPAPGSTA 16  
Db 1 GVTAPDTRPAPGSTA 16  
  
RESULT 37  
ADD88878  
ID ADD88878 standard; peptide; 20 AA.  
XX  
AC ADD88878;

XX 29-JAN-2004 (first entry)  
DT Human MUC1 repeat sequence SEQ ID NO:10.  
XX  
DE Human MUC1 repeat sequence SEQ ID NO:10.  
XX  
KW glycolipopeptide; cytostatic; antibacterial; virucide; antiparasitic; vaccine; gene therapy; immunotherapeutic; cancer; cross-reactive epitope; immune response; parasite; MUC1.  
XX  
OS Homo sapiens.  
XX WO2003089574-A2.  
XX  
PD 30-OCT-2003.  
XX  
PF 09-APR-2003; 2003WO-US010750.  
XX  
PR 15-APR-2002; 2002US-0372105P.  
PR 06-MAY-2002; 2002US-0377595P.  
XX  
PA (BIOM-) BIOMIRA INC.  
XX  
PI Koganty RR, Jiang Z, Yalamati D, Gandhi S, Budzynski W;  
PI Krantz MJ, Longenecker BM;  
XX  
DR WPI; 2003-865440/80.  
XX  
PT New glycolipopeptide comprising a disease-associated epitope or a lipidated interior amino acid, useful as an immunotherapeutic, as a vaccine against cancers and pathogens (e.g. virus or bacteria), or as diagnostic reagents.  
XX  
PS Claim 11; SEQ ID NO 10; 167pp; English.  
XX  
CC The invention relates to a novel non-naturally occurring glycolipopeptide comprising at least 5 amino acids, at least one amino acid being a glycosylated amino acid and at least one amino acid being a lipidated amino acid, where at least one lipidated amino acid is an interior amino acid, the glycolipopeptide comprising at least one disease-associated epitope. A peptide of the invention has cytostatic, antibacterial, virucide, and antiparasitic activity, and may have a use as a vaccine, and in gene therapy. The glycolipopeptide is useful as an immunotherapeutic or as a vaccine against cancers and pathogens (e.g. microbes, toxins, parasites or viruses) presenting cross-reactive epitopes, or as diagnostic reagents. The lipidated amino acid is useful for specific modulation of immune responses to an antigen. The present sequence is used in the exemplification of the invention.  
XX  
SQ Sequence 20 AA;  
  
Query Match 100.0%; Score 83; DB 7; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GVTAPDTRPAPGSTA 16  
Db 1 GVTAPDTRPAPGSTA 16  
  
RESULT 38  
AAW21975  
ID AAW21975 standard; peptide; 21 AA.  
XX  
AC AAW21975;  
XX  
DT 30-OCT-1997 (first entry)  
XX  
DE Human MUC1 VNTR peptide Muc pep1.  
XX  
KW Cancer; vaccine; peptide mimic; mucin; MUC1; Gal alpha(1,3)Gal;  
KW immunotherapy; therapy; adenocarcinoma; VNTR;  
XX variable number of tandem repeat.  
XX

OS Homo sapiens.  
 XX WO97111715-A1.  
 PN XX  
 XX 03-APR-1997.  
 PD XX  
 XX 27-SEP-1996; 96WO-AU000617.  
 PF XX  
 XX 27-SEP-1995; 95AU-00005680.  
 PR XX  
 XX (AUST-) AUSTIN RES INST.  
 PA XX  
 XX Sandrin MS, Mckenzie IFC, Apostolopoulos V;  
 PI XX  
 XX WPI; 1997-212670/19.  
 DR XX  
 XX Cancer vaccine containing MUC1 peptide mimic - used in the treatment of  
 PT adenocarcinoma.  
 PT XX  
 XX Example 1; Page 27; 55pp; English.  
 PS XX  
 XX Muc pep1 is a peptide derived from the variable number of tandem repeat  
 CC (VNTR) region (residues 13-32) of human cancer mucin MUC1. It is capable  
 CC of binding to anti-Gal alpha(1,3)Gal antibody. A novel cancer vaccine  
 CC comprises a peptide (AAW21680-86) which mimics MUC1 or other cancer  
 CC peptides and one or more pharmaceutically acceptable carrier or diluent,  
 CC optionally in association with an appropriate carrier peptide or other  
 CC therapeutic agent. Mucin VNTR peptides (AAW21975-85) may also be used.  
 CC The vaccine is used in the treatment of a patient suffering from, or with  
 CC a predisposition to, adenocarcinoma (claimed)  
 CC XX  
 SQ Sequence 21 AA;  
 Query Match 100.0%; Score 83; DB 2; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSA PDTRPAGSTA 16  
 DB 5 GVTSA PDTRPAGSTA 20  
 RESULT 39  
 AAW22249  
 ID AAW22249 standard; peptide; 21 AA.  
 XX  
 AC AAW22249;  
 XX  
 XX 03-DEC-1997 (first entry)  
 DT  
 DE Muc pep 1, anti-Gal-alpha(1,3)Gal antibody binding peptide.  
 XX  
 KW Consensus sequence; antibody binding peptide; xenotransplant rejection;  
 KW cell surface protein; anti-Gal-alpha(1,3)Gal antibody; Gal-alpha(1,3)Gal;  
 KW endothelial cell; complement cascade; coagulation cascade; immunogen;  
 KW hyperacute rejection; human; xenograft organ; therapy; mouse.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO97111963-A1.  
 PN  
 XX 03-APR-1997.  
 PD  
 XX 27-SEP-1996; 96WO-AU000616.  
 PF  
 XX 27-SEP-1995; 95AU-00005680.  
 PR  
 XX (AUST-) AUSTIN RES INST.  
 PA  
 XX Sandrin MS, Mckenzie IFC, Gallop MA;  
 PI  
 XX WPI; 1997-212848/19.  
 DR  
 XX

PT Peptide capable of binding to anti-Gal-alpha(1,3)Gal antibodies - useful  
 PT for preventing hyper-acute xeno-transplant rejection in humans.  
 XX  
 PS Example 2; Page 26; 49pp; English.  
 XX  
 CC AAW22245-W22260 represent examples of peptides of the invention. The  
 CC peptides of the invention are capable of binding to anti-Gal-  
 CC alpha(1,3)Gal (galactose in an alpha(1,3) linkage with galactose)  
 CC antibodies. The peptides of the invention contain a consensus sequence  
 CC represented by one of the peptides shown in AAW22238-W22244. The binding  
 CC of the anti-Gal-alpha(1,3)Gal antibody to the Gal-alpha(1,3)Gal epitope  
 CC on the surface of endothelial cells results in the activation of both the  
 CC complement and coagulation cascade, resulting in hyperacute rejection.  
 CC Higher primates do not make Gal-alpha(1,3)Gal but humans do possess  
 CC natural antibodies to Gal-alpha(1,3)Gal(15,18). The binding peptides when  
 CC in association with other immunoassay components are useful for the  
 CC detection of anti-Gal-alpha(1,3)Gal antibodies in human biological  
 CC samples. The binding peptides are useful for blocking immunogenic  
 CC activity of human anti-Gal-alpha(1,3)Gal antibodies. They are also useful  
 CC for preventing hyperacute xenotransplant rejection in humans by  
 CC administration of the peptide to the patient who is to receive a donor  
 CC animal organ, in particular the peptide is administered as a vaccine and  
 CC induces immunological tolerance to the Gal-alpha(1,3)Gal epitope. Solid  
 CC substrates which have the binding peptide immobilised to them, are useful  
 CC for the removal of anti-Gal(1,3) antibodies from plasma of human patients  
 CC who will receive a xenograft organ, by passing the plasma over the  
 CC substrate and then returning the plasma to the patient  
 CC  
 SQ Sequence 21 AA;  
 Query Match 100.0%; Score 83; DB 2; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSA PDTRPAGSTA 16  
 DB 5 GVTSA PDTRPAGSTA 20  
 RESULT 40  
 AAY25111  
 ID AAY25111 standard; peptide; 21 AA.  
 XX  
 AC AAY25111;  
 XX  
 XX 25-AUG-1999 (first entry)  
 DT  
 DE Synthetic glycopeptide #1 homologous to MUC1.  
 XX  
 XX Immunodominant; epitope; MUC1; epithelial mucin; tumour; vaccine;  
 KW mammary carcinoma; colorectal carcinoma; pancreas carcinoma.  
 KW  
 XX Synthetic.  
 OS  
 XX DE19758400-A1.  
 PN  
 XX 01-JUL-1999.  
 PD  
 XX 30-DEC-1997; 97DE-01058400.  
 PF  
 XX 30-DEC-1997; 97DE-01058400.  
 PR  
 XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 XX (HANI/) HANISCH F.  
 PA  
 XX Karsten U, Hanisch F, Paulsen H;  
 PI  
 XX WPI; 1999-372374/32.  
 DR  
 XX New tumour vaccine - for treating tumour cells of mammary tumours,  
 PT colorectal and pancreas carcinomas.  
 PT  
 XX Example 1; Page 3; 6pp; German.  
 PS

XX This invention describes a novel tumour vaccine which comprises a  
CC synthetic peptide derived from human epithelial mucin MUC1. This active  
CC tumour vaccine can be used against tumour cells from mammary, colorectal  
CC or pancreas carcinomas. This sequence represents a synthetic glycopeptide  
CC which is homologous to an immunodominant epitope from human MUC1  
XX  
SQ Sequence 21 AA;

Query Match 100.0%; Score 83; DB 2; Length 21;  
Best Local Similarity 100.0%; Pred. No. 6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||  
DB 3 GVTSAPDTRPAPGSTA 18

Search completed: May 6, 2004, 16:29:07  
Job time : 42.4359 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:27:19 ; Search time 13.5385 Seconds  
(without alignments)  
61.013 Million cell updates/sec

Title: US-10-070-566-6  
Perfect score: 83  
Sequence: 1 GVTSAPDTRPAPGSTA 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	83	100.0	16	3 US-09-339-944-7	Sequence 7, Appli
2	83	100.0	16	3 US-08-737-896-1	Sequence 1, Appli
3	83	100.0	16	4 US-09-497-232-15	Sequence 15, Appl
4	83	100.0	16	4 US-09-651-265-12	Sequence 12, Appl
5	83	100.0	16	5 PCT-US96-09951-1	Sequence 1, Appli
6	83	100.0	20	2 US-08-833-807-8	Sequence 8, Appli
7	83	100.0	20	3 US-09-339-944-6	Sequence 6, Appli
8	83	100.0	20	3 US-09-223-043-8	Sequence 8, Appli
9	83	100.0	20	3 US-09-291-351-1	Sequence 1, Appli
10	83	100.0	20	4 US-09-043-731-16	Sequence 16, Appl
11	83	100.0	20	4 US-09-593-870A-20	Sequence 20, Appl
12	83	100.0	20	4 US-09-497-232-10	Sequence 10, Appl
13	83	100.0	20	4 US-09-651-265-6	Sequence 6, Appli
14	83	100.0	21	2 US-08-833-807-7	Sequence 7, Appli
15	83	100.0	21	3 US-09-223-043-7	Sequence 7, Appli
16	83	100.0	21	4 US-09-043-731-15	Sequence 15, Appl
17	83	100.0	21	4 US-09-593-870A-19	Sequence 19, Appl
18	83	100.0	24	3 US-08-737-896-5	Sequence 5, Appli
19	83	100.0	24	4 US-09-497-232-23	Sequence 23, Appl
20	83	100.0	24	5 PCT-US96-09951-5	Sequence 5, Appli
21	83	100.0	25	2 US-08-288-059-28	Sequence 28, Appl
22	83	100.0	25	4 US-09-217-306B-3	Sequence 3, Appli
23	83	100.0	25	4 US-09-497-232-5	Sequence 5, Appli
24	83	100.0	30	3 US-08-737-896-6	Sequence 6, Appli
25	83	100.0	30	3 US-08-134-198E-13	Sequence 13, Appl
26	83	100.0	30	5 PCT-US96-09951-6	Sequence 6, Appli
27	83	100.0	32	4 US-09-593-870A-46	Sequence 46, Appl

28	83	100.0	40	1 US-08-099-354-1	Sequence 1, Appli
29	83	100.0	40	2 US-08-288-059-7	Sequence 7, Appli
30	83	100.0	134	4 US-09-646-028-1	Sequence 1, Appli
31	83	100.0	137	4 US-09-646-028-2	Sequence 2, Appli
32	83	100.0	138	4 US-09-646-028-3	Sequence 3, Appli
33	83	100.0	156	4 US-09-646-028-4	Sequence 4, Appli
34	83	100.0	172	4 US-09-646-028-49	Sequence 49, Appl
35	83	100.0	177	4 US-09-646-028-54	Sequence 54, Appl
36	79	95.2	20	3 US-08-134-198E-35	Sequence 35, Appl
37	79	95.2	30	4 US-09-593-870A-47	Sequence 47, Appl
38	77	92.8	20	1 US-08-328-536-1	Sequence 1, Appli
39	77	92.8	28	2 US-08-488-161-9	Sequence 9, Appli
40	77	92.8	28	3 US-09-273-685-9	Sequence 9, Appli
41	77	92.8	28	5 PCT-US95-11934-9	Sequence 9, Appli
42	75	90.4	24	4 US-09-217-306B-16	Sequence 16, Appl
43	75	90.4	24	4 US-09-217-306B-17	Sequence 17, Appl
44	74	89.2	1867	2 US-08-479-537A-5	Sequence 5, Appli
45	74	89.2	1867	3 US-09-083-116-5	Sequence 5, Appli

## ALIGNMENTS

RESULT 1  
US-09-339-944-7  
; Sequence 7, Application US/09339944  
; Patent No. 6114129  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, Babita  
; APPLICANT: LONGENECKER, B. Michael  
; TITLE OF INVENTION: METHODS OF DETECTING T-CELL ACTIVATION AND TREATING  
; TITLE OF INVENTION: DISORDERS ASSOCIATED WITH T-CELL DYSFUNCTION  
; FILE REFERENCE: 042881/0129  
; CURRENT APPLICATION NUMBER: US/09/339,944  
; CURRENT FILING DATE: 1999-06-25  
; EARLIER APPLICATION NUMBER: 60/090,916  
; EARLIER FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-339-944-7

Query Match 100.0%; Score 83; DB 3; Length 16;  
Best local Similarity 100.0%; Pred. No. 4.4e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||||  
Db 1 GVTSAPDTRPAPGSTA 16

RESULT 2  
US-08-737-896-1  
; Sequence 1, Application US/08737896  
; Patent No. 616804  
; GENERAL INFORMATION:  
; APPLICANT: Samuel, John  
; APPLICANT: Kwon, Glen S.  
; TITLE OF INVENTION: METHOD FOR ELICITING THI-SPECIFIC  
; TITLE OF INVENTION: IMMUNE RESPONSE  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette



```

; REFERENCE/DOCKET NUMBER: 042881/0114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
;
US-09-497-232-15
;
Query Match 100.0%; Score 83; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPARGSTA 16
Db 1 GVTSAPDTRPARGSTA 16
|||||

RESULT 4
US-09-651-265-12
; Sequence 12, Application US/09651265
; Patent No. 6602660
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, B. MICHAEL
; TITLE OF INVENTION: METHODS OF DETECTING T-CELL ACTIVATION
; FILE REFERENCE: 042881/0151
; CURRENT APPLICATION NUMBER: US/09/651,265
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 09/339,344
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/090,916
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-651-265-12
;
Query Match 100.0%; Score 83; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPARGSTA 16
Db 1 GVTSAPDTRPARGSTA 16
|||||

RESULT 5
PCT-US96-09951-1
; Sequence 1, Application PC/TUS9609951
; GENERAL INFORMATION:
; APPLICANT: The Governors of the University of Alberta
; TITLE OF INVENTION: A METHOD FOR ELICITING A TH1-SPECIFIC
; TITLE OF INVENTION: IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037

```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09951
; FILING DATE: 06-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Learn, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07254/037W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SP1-007 MUC1
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..16
PCT-US96-09951-1

Query Match 100.0%; Score 83; DB 5; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSA PDTRPAGSTA 16
Db 1 GVTSA PDTRPAGSTA 16

RESULT 6
US-08-833-807-8
; Sequence 8, Application US/08833807
; Patent No. 5989552
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann Dorfman Herrell and Skillman
; STREET: Suite 720, 1601 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,807
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/340,711
; FILING DATE: 16-NOV-1994
; APPLICATION NUMBER: AU PM3223
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
```

```
;
; REGISTRATION NUMBER: 27,643
; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215)563-4100
; TELEFAX: (215)563-4044
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-833-807-8

Query Match 100.0%; Score 83; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSA PDTRPAGSTA 16
Db 4 GVTSA PDTRPAGSTA 19

RESULT 7
US-09-339-944-6
; Sequence 6, Application US/09339944
; Patent No. 6114129
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; APPLICANT: LONGENECKER, B. Michael
; TITLE OF INVENTION: METHODS OF DETECTING T-CELL ACTIVATION AND TREATING
; TITLE OF INVENTION: DISORDERS ASSOCIATED WITH T-CELL DYSFUNCTION
; FILE REFERENCE: 042881/0129
; CURRENT APPLICATION NUMBER: US/09/339,944
; CURRENT FILING DATE: 1999-06-25
; EARLIER APPLICATION NUMBER: 60/090,916
; EARLIER FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-339-944-6

Query Match 100.0%; Score 83; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSA PDTRPAGSTA 16
Db 1 GVTSA PDTRPAGSTA 16

RESULT 8
US-09-223-043-8
; Sequence 8, Application US/09223043
; Patent No. 6177256
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann Dorfman Herrell and Skillman
; STREET: Suite 720, 1601 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/223,043  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/833,807  
FILING DATE:  
APPLICATION NUMBER: AU PM3223  
FILING DATE: 24-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hagan, Patrick J.  
REGISTRATION NUMBER: 27,643  
REFERENCE/DOCKET NUMBER: 530547/PAS/MKR  
TELEPHONE: (215)563-4100  
TELEFAX: (215)563-4044  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-223-043-8

Query Match 100.0%; Score 83; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSPDTRPAPGSTA 16  
|||  
Db 4 GVTSPDTRPAPGSTA 19

## RESULT 9

US-09-291-351-1  
Sequence 1, Application US/09291351  
Patent No. 6281004  
GENERAL INFORMATION:  
APPLICANT: Bogen, Steven A.  
APPLICANT: Radcliffe, Gail E.  
TITLE OF INVENTION: Quality Control for Cytochemical Assays  
FILE REFERENCE: CYL98-03  
CURRENT APPLICATION NUMBER: US/09/291,351  
CURRENT FILING DATE: 1999-04-14  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Human  
US-09-291-351-1

Query Match 100.0%; Score 83; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSPDTRPAPGSTA 16  
|||  
Db 4 GVTSPDTRPAPGSTA 19

## RESULT 10

US-09-043-731-16  
Sequence 16, Application US/09043731A  
Patent No. 6344203  
GENERAL INFORMATION:  
APPLICANT: The Austin Research Institute  
TITLE OF INVENTION: Mimicking Peptides in Cancer Therapy

FILE REFERENCE: CALA-200  
CURRENT APPLICATION NUMBER: US/09/043,731A  
CURRENT FILING DATE: 1998-06-23  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 16  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: single  
OTHER INFORMATION: stranded linear peptide  
US-09-043-731-16

Query Match 100.0%; Score 83; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSPDTRPAPGSTA 16  
|||  
Db 4 GVTSPDTRPAPGSTA 19

## RESULT 11

US-09-593-870A-20  
Sequence 20, Application US/09593870A  
Patent No. 6548643  
GENERAL INFORMATION:  
APPLICANT: McKenzie, Ian F.C.  
APPLICANT: Apostolopoulos, Vasso  
APPLICANT: Pletersz, Geoff Allan  
TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their  
TITLE OF INVENTION: Use in Immunotherapy  
FILE REFERENCE: 2368-McKenzie  
CURRENT APPLICATION NUMBER: US/09/593,870A  
CURRENT FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: 09/223,043  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 20  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-593-870A-20

Query Match 100.0%; Score 83; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSPDTRPAPGSTA 16  
|||  
Db 4 GVTSPDTRPAPGSTA 19

## RESULT 12

US-09-497-232-10  
Sequence 10, Application US/09497232  
Patent No. 6600012  
GENERAL INFORMATION:  
APPLICANT: AGRAWAL, Babita  
KRANTZ, Mark J.  
REDDISH, Mark A.  
LONGENECKER, B. Michael  
TITLE OF INVENTION: METHOD FOR GENERATING ACTIVATED T-CELLS  
AND ANTIGEN-PULSED ANTIGEN-PRESENTING CELLS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSER: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.

```

; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0; Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/497,232
; FILING DATE: 03-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,410
; FILING DATE: 08-MAY-1998
; APPLICATION NUMBER: US 60/045,949
; FILING DATE: 08-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 042881/0114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-497-232-10

```

```

Query Match      100.0%; Score 83; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GVTSPDTRPAGSTA 16
      |||||
Db      1 GVTSPDTRPAGSTA 16

```

```

RESULT 13
US-09-651-265-6
; Sequence 6, Application US/09651265
; Patent No. 6502560
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, B. MICHAEL
; TITLE OF INVENTION: METHODS OF DETECTING T-CELL ACTIVATION
; FILE REFERENCE: 042881/0151
; CURRENT APPLICATION NUMBER: US/09/651,265
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 09/339,344
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/090,916
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-651-265-6

```

```

Query Match      100.0%; Score 83; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GVTSPDTRPAGSTA 16
      |||||

```

```

Db      1 GVTSPDTRPAGSTA 16

RESULT 14
US-08-833-807-7
; Sequence 7, Application US/08833807
; Patent No. 5989552
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Iann Dorfman Herrell and Skillman
; STREET: Suite 720, 1601 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,807
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/340,711
; FILING DATE: 16-NOV-1994
; APPLICATION NUMBER: AU PM3223
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215)563-4100
; TELEFAX: (215)563-4044
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-833-807-7

```

```

Query Match      100.0%; Score 83; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GVTSPDTRPAGSTA 16
      |||||
Db      5 GVTSPDTRPAGSTA 20

```

```

RESULT 15
US-09-223-043-7
; Sequence 7, Application US/09223043
; Patent No. 6177256
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Iann Dorfman Herrell and Skillman

```

STREET: Suite 720, 1601 Market Street  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: United States of America  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/223,043  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/833,807  
FILING DATE:  
APPLICATION NUMBER: AU PM3223  
FILING DATE: 24-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hagan, Patrick J.  
REGISTRATION NUMBER: 27,643  
REFERENCE/DOCKET NUMBER: 530547/PAS/MKR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215)563-4300  
TELEFAX: (215)563-4044  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-223-043-7

Query Match 100.0%; Score 83; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.8e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
Db 5 GVTSAPDTRPAPGSTA 20

RESULT 16  
US-09-043-731-15  
Sequence 15, Application US/09043731A  
Patent No. 6344203  
GENERAL INFORMATION:  
APPLICANT: The Austin Research Institute  
TITLE OF INVENTION: Mimicking Peptides in Cancer Therapy  
FILE REFERENCE: CALA-200  
CURRENT APPLICATION NUMBER: US/09/043,731A  
CURRENT FILING DATE: 1998-06-23  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 15  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: single  
OTHER INFORMATION: stranded linear peptide  
US-09-043-731-15

Query Match 100.0%; Score 83; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.8e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
Db 5 GVTSAPDTRPAPGSTA 20

RESULT 17  
US-09-593-870A-19  
Sequence 19, Application US/09593870A  
Patent No. 6548643  
GENERAL INFORMATION:  
APPLICANT: McKenzie, Ian F.C.  
APPLICANT: Apostolopoulos, Vasso  
APPLICANT: Pietersz, Geoff Allan  
TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their  
TITLE OF INVENTION: Use in Immunotherapy  
FILE REFERENCE: 2368-McKenzie  
CURRENT APPLICATION NUMBER: US/09/593,870A  
CURRENT FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: 09/223,043  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 19  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-593-870A-19

Query Match 100.0%; Score 83; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.8e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
Db 5 GVTSAPDTRPAPGSTA 20

RESULT 18  
US-08-737-896-5  
Sequence 5, Application US/08737896  
Patent No. 6168804  
GENERAL INFORMATION:  
APPLICANT: Samuel, John  
APPLICANT: Kwon, Glen S.  
TITLE OF INVENTION: METHOD FOR ELICITING THI-SPECIFIC  
TITLE OF INVENTION: IMMUNE RESPONSE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/737,896  
FILING DATE: 24-SEP-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,499  
FILING DATE: 07-JUN-1996  
APPLICATION NUMBER: PCT/US96/09551  
FILING DATE: 07-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07254/037001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SPQ-065 MUC1
US-08-737-896-5

Query Match      100.0%; Score 83; DB 3; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GVTSA PDTRPAGSTA 16
Db      7 GVTSA PDTRPAGSTA 22

RESULT 19
US-09-497-232-23
; Sequence 23, Application US/09497232
; Patent No. 6600012
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; REDDISH, Mark J.
; LONGENECKER, B. Michael
; TITLE OF INVENTION: METHOD FOR GENERATING ACTIVATED T-CELLS
; AND ANTIGEN-PULSED ANTIGEN-PRESENTING CELLS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/497,232
; FILING DATE: 03-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,410
; FILING DATE: 08-MAY-1998
; APPLICATION NUMBER: US 60/045,949
; FILING DATE: 08-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 042881/0114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-497-232-23

Query Match      100.0%; Score 83; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GVTSA PDTRPAGSTA 16
Db      7 GVTSA PDTRPAGSTA 22
```

```
RESULT 20
PCT-US96-09951-5
; Sequence 5, Application PC/TUS9609951
; GENERAL INFORMATION:
; APPLICANT: The Governors of the University of Alberta
; TITLE OF INVENTION: A METHOD FOR ELICITING A Th1-SPECIFIC
; TITLE OF INVENTION: IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09951
; FILING DATE: 06-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Leairn, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07254/037W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SPQ-065 MUC1
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..24
PCT-US96-09951-5

Query Match      100.0%; Score 83; DB 5; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GVTSA PDTRPAGSTA 16
Db      7 GVTSA PDTRPAGSTA 22

RESULT 21
US-08-288-059-28
; Sequence 28, Application US/08288059
; Patent No. 5827666
; GENERAL INFORMATION:
; APPLICANT: FINN, OLIVERA J.
; APPLICANT: FONTENOT, J. D.
; APPLICANT: MONTEJARO, RONALD C.
; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
```



```
;
;
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,059
; FILING DATE: 08-AUG-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: CHAPIN, MARLANA K.
; REGISTRATION NUMBER: 35,843
; REFERENCE/DOCKET NUMBER: 61137/205204
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-288-059-28

Query Match 100.0%; Score 83; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16
Db 1 GVTSAPDTRPAPGSTA 16

RESULT 22
US-09-217-306B-3
; Sequence 3, Application US/09217306B
; Patent No. 6465220
; GENERAL INFORMATION:
; APPLICANT: Hassan, Helle
; APPLICANT: Clausen, Henrik
; APPLICANT: Bennett, Eric P.
; TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase
; FILE REFERENCE: 8850*1
; CURRENT APPLICATION NUMBER: US/09/217,306B
; CURRENT FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: REPEAT
; LOCATION: (1)..(25)
; OTHER INFORMATION: MUC-1 tandem repeat
;
US-09-217-306B-3

Query Match 100.0%; Score 83; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16
Db 7 GVTSAPDTRPAPGSTA 22

RESULT 23
US-09-497-232-5
; Sequence 5, Application US/09497232
; Patent No. 6600012
```

```
;
;
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; KRANTZ, Mark J.
; REDDISH, Mark A.
; LONGENECKER, B. Michael
; TITLE OF INVENTION: METHOD FOR GENERATING ACTIVATED T-CELLS
; AND ANTIGEN-PULSED ANTIGEN-PRESENTING CELLS
;
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/497,232
; FILING DATE: 03-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,410
; FILING DATE: 08-MAY-1998
; APPLICATION NUMBER: US 60/045,949
; FILING DATE: 08-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 042881/0114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
;
US-09-497-232-5

Query Match 100.0%; Score 83; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16
Db 8 GVTSAPDTRPAPGSTA 23

RESULT 24
US-08-737-896-6
; Sequence 6, Application US/08737896
; Patent No. 6168804
; GENERAL INFORMATION:
; APPLICANT: Samuel, John
; APPLICANT: Kwon, Glen S.
; TITLE OF INVENTION: METHOD FOR ELICITING THI-SPECIFIC
; IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,896
; FILING DATE: 24-SEP-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,499
; FILING DATE: 07-JUN-1996
; APPLICATION NUMBER: PCT/US96/09551
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07254/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SPI-070 MUC1
;
US-08-737-896-6

```

```

Query Match 100.0%; Score 83; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GVTSPDTRPAPGSTA 16
Db 13 GVTSPDTRPAPGSTA 28

```

```

RESULT 25
US-08-134-198E-13
; Sequence 13, Application US/08134198E
; Patent No. 619085
; GENERAL INFORMATION:
; APPLICANT: CANCER RESEARCH FUND
; APPLICANT: OF CONTRA COSTA
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: FUSION PROTEIN CONTAINING HMEF
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower Street, Suite 1900
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/134,198E
; FILING DATE: October 8, 1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P65 38208 (CRFC-003C)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 622-7700
; TELEFAX: (213) 489-4210

```

```

; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-134-198E-13

```

```

Query Match 100.0%; Score 83; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GVTSPDTRPAPGSTA 16
Db 1 GVTSPDTRPAPGSTA 16

```

```

RESULT 26
PCT-US96-09951-6
; Sequence 6, Application PC/TUS9609951
; GENERAL INFORMATION:
; APPLICANT: The Governors of the University of Alberta
; TITLE OF INVENTION: A METHOD FOR ELICITING A TH1-SPECIFIC
; TITLE OF INVENTION: IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09951
; FILING DATE: 06-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Learn, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07254/037W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SPI-070 MUC1
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..30
;
PCT-US96-09951-6

```

```

Query Match 100.0%; Score 83; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GVTSPDTRPAPGSTA 16
Db 13 GVTSPDTRPAPGSTA 28

```

RESULT 27

US-09-593-870A-46  
 ; Sequence 46, Application US/09593870A  
 ; Patent No. 6548643  
 ; GENERAL INFORMATION:  
 ; APPLICANT: McKenzie, Ian F.C.  
 ; APPLICANT: Apostolopoulos, Vasso  
 ; APPLICANT: Pietersz, Geoff Allan  
 ; TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their  
 ; FILE REFERENCE: 2368-McKenzie  
 ; CURRENT APPLICATION NUMBER: US/09/593,870A  
 ; CURRENT FILING DATE: 2000-06-14  
 ; PRIOR APPLICATION NUMBER: 09/223,043  
 ; PRIOR FILING DATE: 1998-12-30  
 ; NUMBER OF SEQ ID NOS: 69  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 46  
 ; LENGTH: 32  
 ; TYPE: PRCT  
 ; ORGANISM: Homo sapiens  
 US-09-593-870A-46

Query Match 100.0%; Score 83; DB 4; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 9e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
 Db 16 GVTSAPDTRPAPGSTA 31

RESULT 28  
 US-08-099-354-1  
 ; Sequence 1, Application US/08099354  
 ; Patent No. 5744144  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FINN, OLIVERA J.  
 ; APPLICANT: FONTENOT, J. D.  
 ; APPLICANT: MONTELARO, RONALD C.  
 ; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN  
 ; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
 ; STREET: 1100 NEW YORK AVENUE, N.W.  
 ; CITY: WASHINGTON  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/099,354  
 ; FILING DATE: 30-JUL-1993  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: SIRILLA, GEORGE M.  
 ; REGISTRATION NUMBER: 18221  
 ; REFERENCE/DOCKET NUMBER: 6137/202246  
 ; TELEPHONE: 202-861-3536  
 ; TELEFAX: 202-822-0944  
 ; TELEX: 6714627 CUSH  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 40 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide

US-08-099-354-1  
 Query Match 100.0%; Score 83; DB 1; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 Db 17 GVTSAPDTRPAPGSTA 32

RESULT 29  
 US-08-288-059-7  
 ; Sequence 7, Application US/08288059  
 ; Patent No. 5827656  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FINN, OLIVERA J.  
 ; APPLICANT: FONTENOT, J. D.  
 ; APPLICANT: MONTELARO, RONALD C.  
 ; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN  
 ; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 36  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
 ; STREET: 1100 NEW YORK AVENUE, N.W.  
 ; CITY: WASHINGTON  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/288,059  
 ; FILING DATE: 08-AUG-1994  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CHAPIN, MARLANA K.  
 ; REGISTRATION NUMBER: 35,843  
 ; REFERENCE/DOCKET NUMBER: 61137/205204  
 ; TELEPHONE: 202-861-3711  
 ; TELEFAX: 202-822-0944  
 ; TELEX: 6714627 CUSH  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 40 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide

US-08-288-059-7  
 Query Match 100.0%; Score 83; DB 2; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GVTSAPDTRPAPGSTA 16  
 Db 17 GVTSAPDTRPAPGSTA 32

RESULT 30  
 US-09-646-028-1  
 ; Sequence 1, Application US/09646028  
 ; Patent No. 6562347  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kwak, Larry  
 ; APPLICANT: Biragyn, Arya  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF  
 ; CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES

```
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-1
```

```
Query Match 100.0%; Score 83; DB 4; Length 134;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GVTSPDTRPAPGSTA 16
Db 89 GVTSPDTRPAPGSTA 104
```

# RESULT 31

```
US-09-646-028-2
; Sequence 2, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-2
```

```
Query Match 100.0%; Score 83; DB 4; Length 137;
Best Local Similarity 100.0%; Pred. No. 4.1e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GVTSPDTRPAPGSTA 16
Db 92 GVTSPDTRPAPGSTA 107
```

# RESULT 32

```
US-09-646-028-3
; Sequence 3, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 3
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-3
```

```
Query Match 100.0%; Score 83; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GVTSPDTRPAPGSTA 16
Db 93 GVTSPDTRPAPGSTA 108
```

# RESULT 33

```
US-09-646-028-4
; Sequence 4, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-4
```

```
Query Match 100.0%; Score 83; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 4.7e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GVTSPDTRPAPGSTA 16
Db 111 GVTSPDTRPAPGSTA 126
```

# RESULT 34

```
US-09-646-028-49
; Sequence 49, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-49
```

```
Query Match      100.0%; Score 83; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 5.2e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GVTSAPDTRPAPGSTA 16
      |||||||
Db      81  GVTSAPDTRPAPGSTA 96

RESULT 35
US-09-646-028-54
; Sequence 54, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; FILE REFERENCE: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 54
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-54

Query Match      100.0%; Score 83; DB 4; Length 177;
Best Local Similarity 100.0%; Pred. No. 5.4e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GVTSAPDTRPAPGSTA 16
      |||||||
Db      86  GVTSAPDTRPAPGSTA 101

RESULT 36
US-08-134-198E-35
; Sequence 35, Application US/08134198E
; Patent No. 6190885
; GENERAL INFORMATION:
; APPLICANT: CANCER RESEARCH FUND
; APPLICANT: OF CONTRA COSTA
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: FUSION PROTEIN CONTAINING HMFG
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower Street, Suite 1900
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/134,198E
; FILING DATE: October 8, 1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
```

```
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38208 (CRFC-003C)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 622-7700
; TELEFAX: (213) 489-4210
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-134-198E-35

Query Match      95.2%; Score 79; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GVTSAPDTRPAPGST 15
      |||||||
Db      6  GVTSAPDTRPAPGST 20

RESULT 37
US-09-593-870A-47
; Sequence 47, Application US/09593870A
; Patent No. 6548643
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff Allan
; TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their
; TITLE OF INVENTION: Use in Immunotherapy
; FILE REFERENCE: 2368-McKenzie
; CURRENT APPLICATION NUMBER: US/09/593,870A
; CURRENT FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: 09/223,043
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 47
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-593-870A-47

Query Match      95.2%; Score 79; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GVTSAPDTRPAPGST 15
      |||||||
Db      16  GVTSAPDTRPAPGST 30

RESULT 38
US-08-328-536-1
; Sequence 1, Application US/08328536
; Patent No. 5506343
; GENERAL INFORMATION:
; APPLICANT: Kufe, Donald
; TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR
; TITLE OF INVENTION: CARCINOMA-ASSOCIATED ANTIGENS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
```

```
;
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,536
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/868,352
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-328-536-1

Query Match          92.8%; Score 77; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VTSAPDTRPARGSTA 16
Db      1 VTSAPDTRPARGSTA 15

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,161
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/488,161
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-176
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-488-161-9

RESULT 39
US-08-488-161-9
; Sequence 9, Application US/08488161
; Patent No. 5885577
; GENERAL INFORMATION:
; APPLICANT: Alvarez, Vernon L.
; TITLE OF INVENTION: Antigen Binding Peptides (Abptides) From
; TITLE OF INVENTION: Peptide Libraries
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,161
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-176
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-488-161-9

Query Match          92.8%; Score 77; DB 2; Length 28;
Best Local Similarity 100.0%; Pred. No. 5.9e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VTSAPDTRPARGSTA 16
Db      1 VTSAPDTRPARGSTA 15

;
; RESULT 40
; US-09-273-685-9
; Sequence 9, Application US/09273685
; Patent No. 6015561
; GENERAL INFORMATION:
; APPLICANT: Alvarez, Vernon L.
; TITLE OF INVENTION: Antigen Binding Peptides (Abptides) From
; TITLE OF INVENTION: Peptide Libraries
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/273,685
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/488,161
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-176
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-09-273-685-9

Query Match          92.8%; Score 77; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 5.9e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 VTSAPDTRPARGSTA 16
Db      1 VTSAPDTRPARGSTA 15

;
; Search completed: May 6, 2004, 16:32:37
; Job time : 20.5385 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:31:15 ; Search time 30.359 Seconds  
(without alignments)  
146.285 Million cell updates/sec

Title: US-10-070-566-6  
Perfect score: 83  
Sequence: 1 GVTSAPDTRPAPGSTA 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/JS07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/JS06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83	100.0	16	9	US-09-984-183-17
2	83	100.0	16	9	US-09-984-333-7
3	83	100.0	16	12	US-09-143-379-1
4	83	100.0	18	12	US-09-815-346-3
5	83	100.0	18	14	US-10-106-876-6
6	83	100.0	19	12	US-09-815-346-2
7	83	100.0	19	12	US-09-815-346-6
8	83	100.0	19	14	US-10-106-876-5
9	83	100.0	20	9	US-09-984-183-16
10	83	100.0	20	9	US-09-822-698A-7
11	83	100.0	20	9	US-09-984-333-6
12	83	100.0	20	10	US-09-834-240-1
13	83	100.0	20	12	US-10-296-317-45
14	83	100.0	20	13	US-10-054-488-1
15	83	100.0	20	14	US-10-106-876-19

16	83	100.0	20	14	US-10-292-896-1	Sequence 1, Appli
17	83	100.0	22	14	US-10-106-876-18	Sequence 18, Appl
18	83	100.0	24	14	US-10-292-896-2	Sequence 2, Appli
19	83	100.0	25	10	US-09-881-339-3	Sequence 3, Appli
20	83	100.0	25	15	US-10-417-633-3	Sequence 3, Appli
21	83	100.0	26	12	US-09-815-346-5	Sequence 5, Appli
22	83	100.0	26	14	US-10-106-876-8	Sequence 8, Appli
23	83	100.0	26	14	US-10-106-876-20	Sequence 20, Appl
24	83	100.0	27	12	US-09-815-346-4	Sequence 4, Appli
25	83	100.0	27	12	US-09-870-691-1	Sequence 1, Appli
26	83	100.0	27	12	US-09-939-191-1	Sequence 1, Appli
27	83	100.0	27	14	US-10-106-876-7	Sequence 7, Appli
28	83	100.0	27	14	US-10-106-876-9	Sequence 9, Appli
29	83	100.0	27	14	US-10-106-876-12	Sequence 12, Appl
30	83	100.0	28	14	US-10-106-876-13	Sequence 13, Appl
31	83	100.0	31	9	US-09-984-183-18	Sequence 18, Appl
32	83	100.0	31	9	US-09-984-183-19	Sequence 19, Appl
33	83	100.0	36	12	US-10-296-317-64	Sequence 64, Appl
34	83	100.0	38	12	US-10-296-317-56	Sequence 56, Appl
35	83	100.0	40	10	US-09-996-069-9	Sequence 9, Appli
36	83	100.0	43	14	US-10-106-876-14	Sequence 14, Appl
37	83	100.0	46	10	US-09-996-069-5	Sequence 5, Appli
38	83	100.0	46	14	US-10-106-876-1	Sequence 1, Appli
39	83	100.0	50	12	US-10-449-831A-5	Sequence 5, Appli
40	83	100.0	100	9	US-09-985-131-6	Sequence 6, Appli
41	83	100.0	134	14	US-10-335-394-1	Sequence 1, Appli
42	83	100.0	137	14	US-10-335-394-2	Sequence 2, Appli
43	83	100.0	138	14	US-10-335-394-3	Sequence 3, Appli
44	83	100.0	156	14	US-10-335-394-4	Sequence 4, Appli
45	83	100.0	172	14	US-10-335-394-49	Sequence 49, Appl

## ALIGNMENTS

### RESULT 1

US-09-984-183-17  
; Sequence 17, Application US/09984183  
; Patent No. US20020142983A1  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, BABITA  
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE  
; FILE REFERENCES: 042881/0130  
; CURRENT APPLICATION NUMBER: US/09/984,183  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 1998-12-11  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-183-17

Query Match 100.0%; Score 83; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
| | | | | | | | | | | | | | | |  
Db 1 GVTSAPDTRPAPGSTA 16

### RESULT 2

US-09-984-333-7  
; Sequence 7, Application US/09984333  
; Patent No. US20020159969A1  
; GENERAL INFORMATION:

; APPLICANT: AGRAWAL, Babita  
; APPLICANT: LONGENECKER, Bryan Michael  
; APPLICANT: REDDISH, Mark Austin  
; TITLE OF INVENTION: SMALL PEPTIDE-BASED THERAPEUTICS FOR REVERSING  
; TITLE OF INVENTION: CANCER-ASSOCIATED MUC-1 MUCIN-INDUCED IMMUNOSUPPRESSION  
; FILE REFERENCE: 042881/0116  
; CURRENT APPLICATION NUMBER: US/09/984,333  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 09/182,887  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: 60/064,146  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/065,209  
; PRIOR FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-333-7

Query Match 100.0%; Score 83; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||||  
Db 1 GVTSAPDTRPAPGSTA 16

## RESULT 3

US-09-143-379-1  
; Sequence 1, Application US/09143379  
; Publication No. US20040077826A1  
; GENERAL INFORMATION:  
; APPLICANT: KOGANTY, R. Rao  
; APPLICANT: QIU, Dongxu  
; APPLICANT: GANDHI, Sham  
; TITLE OF INVENTION: RANDOMLY GENERATED GLYCOPEPTIDE COMBINATORIAL LIBRARIES  
; FILE REFERENCE: 042881/0119  
; CURRENT APPLICATION NUMBER: US/09/143,379  
; PRIOR FILING DATE: 1998-08-28  
; EARLIER APPLICATION NUMBER: 60/056,240  
; EARLIER FILING DATE: 1997-08-28  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Peptide  
; OTHER INFORMATION: derived from cancer-associated MUC1  
US-09-143-379-1

Query Match 100.0%; Score 83; DB 12; Length 16;  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||||  
Db 1 GVTSAPDTRPAPGSTA 16

## RESULT 4

US-09-815-346-3  
; Sequence 3, Application US/09815346  
; Publication No. US20020018806A1  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, BABITA  
; APPLICANT: LONGENECKER, MICHAEL B.  
; APPLICANT: PARKER, JOANNE

; TITLE OF INVENTION: LIPOPEPTIDE ADJUVANTS  
; FILE REFERENCE: 042881/0160  
; CURRENT APPLICATION NUMBER: US/09/815,346  
; CURRENT FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: 60/191,736  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-09-815-346-3

Query Match 100.0%; Score 83; DB 12; Length 18;  
Best Local Similarity 100.0%; Pred. No. 8.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||||  
Db 1 GVTSAPDTRPAPGSTA 16

## RESULT 5

US-10-106-876-6  
; Sequence 6, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-6

Query Match 100.0%; Score 83; DB 14; Length 18;  
Best Local Similarity 100.0%; Pred. No. 8.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||||  
Db 1 GVTSAPDTRPAPGSTA 16

## RESULT 6

US-09-815-346-2  
; Sequence 2, Application US/09815346  
; Publication No. US20020018806A1  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, BABITA  
; APPLICANT: LONGENECKER, MICHAEL B.  
; APPLICANT: PARKER, JOANNE  
; TITLE OF INVENTION: LIPOPEPTIDE ADJUVANTS  
; FILE REFERENCE: 042881/0160  
; CURRENT APPLICATION NUMBER: US/09/815,346



```

; OTHER INFORMATION: peptide of MUC1 protein
US-09-822-698A-7

Query Match      100.0%; Score 83; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAPGSTA 16
Db 4 GVTAPDTRPAPGSTA 19

RESULT 11
US-09-844-333-6
; Sequence 6, Application US/09984333
; Patent No. US20020159969A1
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; APPLICANT: LONGNECKER, Bryan Michael
; APPLICANT: REDDISH, Mark Austin
; TITLE OF INVENTION: SMALL PEPTIDE-BASED THERAPEUTICS FOR REVERSING
; TITLE OF INVENTION: CANCER-ASSOCIATED MUC-1 MUCIN-INDUCED IMMUNOSUPPRESSION
; FILE REFERENCE: 042881/0116
; CURRENT APPLICATION NUMBER: US/09/984,333
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 09/182,887
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 60/064,146
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065,209
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-844-333-6

Query Match      100.0%; Score 83; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAPGSTA 16
Db 1 GVTAPDTRPAPGSTA 16

RESULT 12
US-09-834-240-1
; Sequence 1, Application US/09834240
; Publication No. US20030166261A1
; GENERAL INFORMATION:
; APPLICANT: Sompuram, Seshi R.
; APPLICANT: Ramanathan, Halasya
; TITLE OF INVENTION: Quality Control for Cytochemical Assays
; FILE REFERENCE: 1159.1008-005
; CURRENT APPLICATION NUMBER: US/09/834,240
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: 09/549,855
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 09/291,351
; PRIOR FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-834-240-1

Query Match      100.0%; Score 83; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;

; OTHER INFORMATION: peptide of MUC1 protein
US-09-822-698A-7

Query Match      100.0%; Score 83; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAPGSTA 16
Db 4 GVTAPDTRPAPGSTA 19

RESULT 13
US-10-296-317-45
; Sequence 45, Application US/10296317
; Publication No. US20040057968A1
; GENERAL INFORMATION:
; APPLICANT: CEI-Sci Corp
; APPLICANT: Zimmerman, Daniel S
; APPLICANT: Sarin, Prem S
; TITLE OF INVENTION: T CELL BINDING LIGAND PEPTIDES, PEPTIDE
; FILE REFERENCE: CS-112
; CURRENT APPLICATION NUMBER: US/10/296,317
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/206548
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: PCT/US07/16793
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Muc1 Peptide Mlc
US-10-296-317-45

Query Match      100.0%; Score 83; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAPGSTA 16
Db 1 GVTAPDTRPAPGSTA 16

RESULT 14
US-10-054-488-1
; Sequence 1, Application US/10054488
; Publication No. US20020142047A1
; GENERAL INFORMATION:
; APPLICANT: Mark E. Johnson
; APPLICANT: Tricia Cecil
; APPLICANT: Olivera J. Finn
; TITLE OF INVENTION: MICROSPHERE DELIVERY OF MUCIN PEPTIDES
; FILE REFERENCE: 126.04USU1
; CURRENT APPLICATION NUMBER: US/10/054,488
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/262,699
; PRIOR FILING DATE: 2001-01-19
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-488-1

Query Match      100.0%; Score 83; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAPGSTA 16
Db 1 GVTAPDTRPAPGSTA 16

```

## RESULT 15

US-10-106-876-19  
; Sequence 19, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-19

Query Match 100.0%; Score 83; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTAPDTRPAPGSTA 16  
|||  
Db 1 GVTAPDTRPAPGSTA 16

## RESULT 16

US-10-292-896-1  
; Sequence 1, Application US/10292896  
; Publication No. US20030186850A1  
; GENERAL INFORMATION:  
; APPLICANT: HASSAN, Helle  
; APPLICANT: REIS, Celso A.  
; APPLICANT: BENNETT, Eric P.  
; APPLICANT: CLAUSEN, Henrik  
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GR  
; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS  
; FILE REFERENCE: 4305/1H154-US3  
; CURRENT APPLICATION NUMBER: US/10/292,896  
; PRIOR FILING DATE: 2003-03-17  
; PRIOR APPLICATION NUMBER: US 60/425,204  
; PRIOR FILING DATE: 2002-11-08  
; PRIOR APPLICATION NUMBER: PCT/DK01/00328  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: US 60/203,331  
; PRIOR FILING DATE: 2000-05-11  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-292-896-1

Query Match 100.0%; Score 83; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTAPDTRPAPGSTA 16  
|||  
Db 2 GVTAPDTRPAPGSTA 17

## RESULT 17

US-10-106-876-18  
; Sequence 18, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 18  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-18

Query Match 100.0%; Score 83; DB 14; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTAPDTRPAPGSTA 16  
|||  
Db 1 GVTAPDTRPAPGSTA 16

## RESULT 18

US-10-292-896-2  
; Sequence 2, Application US/10292896  
; Publication No. US20030186850A1  
; GENERAL INFORMATION:  
; APPLICANT: HASSAN, Helle  
; APPLICANT: REIS, Celso A.  
; APPLICANT: BENNETT, Eric P.  
; APPLICANT: CLAUSEN, Henrik  
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GR  
; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS  
; FILE REFERENCE: 4305/1H154-US3  
; CURRENT APPLICATION NUMBER: US/10/292,896  
; PRIOR FILING DATE: 2003-03-17  
; PRIOR APPLICATION NUMBER: US 60/425,204  
; PRIOR FILING DATE: 2002-11-08  
; PRIOR APPLICATION NUMBER: PCT/DK01/00328  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: US 60/203,331  
; PRIOR FILING DATE: 2000-05-11  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-292-896-2

Query Match 100.0%; Score 83; DB 14; Length 24;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTAPDTRPAPGSTA 16  
|||  
Db 7 GVTAPDTRPAPGSTA 22

RESULT 20  
US-10-417-633-3  
; Sequence 3, Application US/10417633  
; Publication No. US20030232399A1  
; GENERAL INFORMATION:  
; APPLICANT: Robertson, John F.R.  
; APPLICANT: Graves, Catharine  
; APPLICANT: Price, Michael R.  
; TITLE OF INVENTION: Cancer Detection Methods and Reagents  
; FILE REFERENCE: 52623-0031 52623-284706  
; CURRENT APPLICATION NUMBER: US/10/417,633  
; CURRENT FILING DATE: 2003-04-16  
; PRIOR APPLICATION NUMBER: US 60/211,886  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 09/881,339  
; PRIOR FILING DATE: 2001-06-14  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic MUC1 peptide TAP2  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (9)..(9)  
; OTHER INFORMATION: T is O-glycosylated with N-acetylgalactosamine  
; FEATURE:  
; NAME/KEY: MISC FEATURE

```

; SEQ ID NO 8
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description
; OTHER INFORMATION: peptide
US-10-106-876-8

```



Query Match 100.0%; Score 83; DB 14; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSPDTRPAPGSTA 16  
Db 8 GVTSPDTRPAPGSTA 23

## RESULT 23

US-10-106-876-20  
; Sequence 20, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 26  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-20

Query Match 100.0%; Score 83; DB 14; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSPDTRPAPGSTA 16  
Db 8 GVTSPDTRPAPGSTA 23

## RESULT 24

US-09-815-346-4  
; Sequence 4, Application US/09815346  
; Publication No. US20020018806A1  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, BABITA  
; APPLICANT: LONGENECKER, MICHAEL B.  
; APPLICANT: PARKER, JOANNE  
; TITLE OF INVENTION: LIPOPEPTIDE ADJUVANTS  
; FILE REFERENCE: 042881/0160  
; CURRENT APPLICATION NUMBER: US/09/815,346  
; PRIOR FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: 60/191,736  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-09-815-346-4

Query Match 100.0%; Score 83; DB 12; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GVTSPDTRPAPGSTA 16  
Db 7 GVTSPDTRPAPGSTA 22

## RESULT 25

US-09-870-691-1  
; Sequence 1, Application US/09870691  
; Publication No. US20020051813A1  
; GENERAL INFORMATION:  
; APPLICANT: BONI, Lawrence  
; APPLICANT: BATENJANY, Michael M.  
; APPLICANT: ROBB, Richard J.  
; APPLICANT: POESCU, Mircea C.  
; APPLICANT: NEVILLE, Mary E.  
; TITLE OF INVENTION: HIGH CHOLESTEROL LIPOMATRIX PREPARATION  
; FILE REFERENCE: 040403/0200  
; CURRENT APPLICATION NUMBER: US/09/870,691  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR FILING DATE: 1999-01-15  
; PRIOR APPLICATION NUMBER: US/09/231,641  
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: US 60/071,688  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: BP1-148  
; OTHER INFORMATION: lipopeptide  
US-09-870-691-1

Query Match 100.0%; Score 83; DB 12; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVTSPDTRPAPGSTA 16  
Db 8 GVTSPDTRPAPGSTA 23

## RESULT 26

US-09-999-191-1  
; Sequence 1, Application US/09999191  
; Publication No. US20020160039A1  
; GENERAL INFORMATION:  
; APPLICANT: BONI, LAWRENCE  
; APPLICANT: WU, FANGJUN  
; APPLICANT: FENNIMORE, ROY  
; APPLICANT: BATENJANY, MICHAEL M.  
; TITLE OF INVENTION: PREPARATION OF LARGE LIPOSOMES BY INFUSION INTO PEG  
; FILE REFERENCE: 042881/0187  
; CURRENT APPLICATION NUMBER: US/09/999,191  
; CURRENT FILING DATE: 2001-12-03  
; PRIOR APPLICATION NUMBER: 60/250,124  
; PRIOR FILING DATE: 2000-12-01  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-09-999-191-1

Query Match 100.0%; Score 83; DB 12; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 100.0%; Score 83; DB 12; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSA PDTRPAGSTA 16  
| | | | | | | | | | | | | | | | | | | | |  
Db 8 GVTSA PDTRPAGSTA 23

RESULT 27  
US-10-106-876-7  
; Sequence 7, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-7

Query Match 100.0%; Score 83; DB 14; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSA PDTRPAGSTA 16  
| | | | | | | | | | | | | | | | | | | | |  
Db 7 GVTSA PDTRPAGSTA 22

RESULT 28  
US-10-106-876-9  
; Sequence 9, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-9

Query Match 100.0%; Score 83; DB 14; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSA PDTRPAGSTA 16  
| | | | | | | | | | | | | | | | | | | | |  
Db 7 GVTSA PDTRPAGSTA 22

RESULT 29  
US-10-106-876-12  
; Sequence 12, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-12

Query Match 100.0%; Score 83; DB 14; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSA PDTRPAGSTA 16  
| | | | | | | | | | | | | | | | | | | | |  
Db 8 GVTSA PDTRPAGSTA 23

RESULT 30  
US-10-106-876-13  
; Sequence 13, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-13

Query Match 100.0%; Score 83; DB 14; Length 28;

```
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAGSTA 16
Db 8 GVTSPDTRPAGSTA 23

RESULT 31
US-09-984-183-18
; Sequence 18, Application US/09984183
; Patent No. US20020142983A1
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, MICHAEL B.
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 042881/0130
; CURRENT APPLICATION NUMBER: US/09/984,183
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/457,354
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/111,973
; PRIOR FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MOD RES
; LOCATION: {17}..(21)
; OTHER INFORMATION: this region may encompass one to five
; OTHER INFORMATION: amino acids selected from cysteine, serine
; OTHER INFORMATION: threonine and methionine
; NAME/KEY: MOD RES
; LOCATION: {25}..(26)
; OTHER INFORMATION: this region may encompass one or two
; OTHER INFORMATION: amino acids selected from cysteine, serine,
; OTHER INFORMATION: threonine and methionine
; NAME/KEY: MOD RES
; LOCATION: {27}..(31)
; OTHER INFORMATION: this region may encompass one to five
; OTHER INFORMATION: amino acids selected from cysteine, serine,
; OTHER INFORMATION: threonine and methionine
US-09-984-183-18

Query Match 100.0%; Score 83; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAGSTA 16
Db 1 GVTSPDTRPAGSTA 16

RESULT 32
US-09-984-183-19
; Sequence 19, Application US/09984183
; Patent No. US20020142983A1
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, MICHAEL B.
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 042881/0130
; CURRENT APPLICATION NUMBER: US/09/984,183
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/457,354
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/111,973
; PRIOR FILING DATE: 1998-12-11
```

```
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 19
LENGTH: 31
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MOD RES
LOCATION: {1}..(5)
OTHER INFORMATION: this region may encompass one to five
OTHER INFORMATION: amino acids selected from cysteine, serine
OTHER INFORMATION: threonine and methionine
NAME/KEY: MOD RES
LOCATION: {9}..(10)
OTHER INFORMATION: this region may encompass one or two
OTHER INFORMATION: amino acids selected from cysteine, serine,
OTHER INFORMATION: threonine and methionine
NAME/KEY: MOD RES
LOCATION: {11}..(15)
OTHER INFORMATION: this region may encompass one to five
OTHER INFORMATION: amino acids selected from cysteine, serine,
OTHER INFORMATION: threonine and methionine
US-09-984-183-19

Query Match 100.0%; Score 83; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAGSTA 16
Db 16 GVTSPDTRPAGSTA 31

RESULT 33
US-10-296-317-64
; Sequence 64, Application US/10296317
; Publication No. US20040057968A1
; GENERAL INFORMATION:
; APPLICANT: Cel-Sci Corp
; APPLICANT: Zimmerman, Daniel S
; APPLICANT: Sarin, Prem S
; TITLE OF INVENTION: T CELL BINDING LIGAND PEPTIDES, PEPTIDE
; FILE REFERENCE: CS-112
; CURRENT APPLICATION NUMBER: US/10/296,317
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/206548
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: PCT/US07/16793
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Peptide construct
US-10-296-317-64

Query Match 100.0%; Score 83; DB 12; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAGSTA 16
Db 17 GVTSPDTRPAGSTA 32

RESULT 34
US-10-296-317-56
; Sequence 56, Application US/10296317
; Publication No. US20040057968A1
; GENERAL INFORMATION:
```

; APPLICANT: CEL-Sci Corp  
; APPLICANT: Zimmerman, Daniel S  
; APPLICANT: Sarin, Prem S  
; TITLE OF INVENTION: T CELL BINDING LIGAND PEPTIDES, PEPTIDE  
; FILE REFERENCE: CS-112  
; CURRENT APPLICATION NUMBER: US/10/296,317  
; CURRENT FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US 60/206548  
; PRIOR FILING DATE: 2000-05-24  
; PRIOR APPLICATION NUMBER: PCT/US07/16793  
; PRIOR FILING DATE: 2001-05-24  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 56  
; LENGTH: 38  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Peptide construct  
US-10-296-317-56

Query Match 100.0%; Score 83; DB 12; Length 38;  
Best Local Similarity 100.0%; Pred. No. 0.00018;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||  
Db 19 GVTSAPDTRPAPGSTA 34

RESULT 35  
US-09-996-069-9  
; Sequence 9, Application US/09996069  
; Publication No. US20030036199A1  
; GENERAL INFORMATION:  
; APPLICANT: Bamdad, Cynthia  
; APPLICANT: Bamdad, R. Shoshana  
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBI  
; TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER  
; FILE REFERENCE: M01015/70071  
; CURRENT APPLICATION NUMBER: US/09/996,069  
; CURRENT FILING DATE: 2001-11-27  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 9  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-996-069-9

Query Match 100.0%; Score 83; DB 10; Length 40;  
Best Local Similarity 100.0%; Pred. No. 0.00019;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||  
Db 16 GVTSAPDTRPAPGSTA 31

RESULT 36  
US-10-106-876-14  
; Sequence 14, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 43  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-10-106-876-14

Query Match 100.0%; Score 83; DB 14; Length 43;  
Best Local Similarity 100.0%; Pred. No. 0.00021;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||  
Db 19 GVTSAPDTRPAPGSTA 34

RESULT 37  
US-09-996-069-5  
; Sequence 5, Application US/09996069  
; Publication No. US20030036199A1  
; GENERAL INFORMATION:  
; APPLICANT: Bamdad, Cynthia  
; APPLICANT: Bamdad, R. Shoshana  
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBI  
; TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER  
; FILE REFERENCE: M01015/70071  
; CURRENT APPLICATION NUMBER: US/09/996,069  
; CURRENT FILING DATE: 2001-11-27  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 46  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-996-069-5

Query Match 100.0%; Score 83; DB 10; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.00022;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
|||  
Db 16 GVTSAPDTRPAPGSTA 31

RESULT 38  
US-10-106-876-1  
; Sequence 1, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL  
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE  
; TITLE OF INVENTION: RESPONSES  
; FILE REFERENCE: 042881-0176  
; CURRENT APPLICATION NUMBER: US/10/106,876  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 60/278,698  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 46  
; TYPE: PRT  
; ORGANISM: Artificial Sequence

```
;
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-106-876-1
```

```
Query Match 100.0%; Score 83; DB 14; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GVTSA PDTRPAPGSTA 16
Db 2 GVTSA PDTRPAPGSTA 17
```

# RESULT 39

```
US-10-449-831A-5
; Sequence 5, Application US/10449831A
; Publication No. US20040029179A1
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; TITLE OF INVENTION: Higher molecular weight entities and uses therefor
; FILE REFERENCE: 2385978
; CURRENT APPLICATION NUMBER: US/10/449,831A
; CURRENT FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: USSN 60/384878
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 237
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: MUC-1 chimeric peptide
US-10-449-831A-5
```

```
Query Match 100.0%; Score 83; DB 12; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GVTSA PDTRPAPGSTA 16
Db 31 GVTSA PDTRPAPGSTA 46
```

# RESULT 40

```
US-09-965-131-6
; Sequence 6, Application US/09965131
; Patent No. US20020160502A1
; GENERAL INFORMATION:
; APPLICANT: Chung, Maureen A.
; APPLICANT: Sharma, Surendra
; APPLICANT: Chang, Helena R.
; APPLICANT: O'Donnell, Mark A.
; TITLE OF INVENTION: RECOMBINANT BCG VACCINES FOR THE
; FILE REFERENCE: WII-014CP
; CURRENT APPLICATION NUMBER: US/09/965,131
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,455
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-131-6
```

```
Query Match 100.0%; Score 83; DB 9; Length 100;
Best Local Similarity 100.0%; Pred. No. 0.00048;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GVTSA PDTRPAPGSTA 16
Db 1 GVTSA PDTRPAPGSTA 16

Search completed: May 6, 2004, 16:40:44
Job time : 31.359 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:26:14 ; Search time 11.0769 Seconds  
(without alignments)  
138.943 Million cell updates/sec

Title: US-10-070-566-6

Perfect score: 83

Sequence: 1 GVTSAPDTRPAPGSTA 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83	100.0	347	2 S10571	mucin 1 precursor,
2	83	100.0	1344	1 A35175	mucin 1 precursor,
3	69	83.1	256	2 A60533	tumor-associated a
4	57	68.7	1334	2 T50568	penicillin binding
5	50	60.2	182	2 S73046	probable D-alanyl-
6	50	60.2	411	2 D86995	probable D-alanyl-
7	49	59.0	78	2 T27876	hypothetical prote
8	49	59.0	175	2 T27875	hypothetical prote
9	49	59.0	447	2 T18264	cellulosome anchor
10	49	59.0	1101	2 G70951	probable ATP-depen
11	47	56.6	633	2 E75277	ABC transporter, A
12	46	55.4	108	2 F72653	hypothetical prote
13	46	55.4	399	2 T18853	probable transcrip
14	45	54.2	115	2 F72570	hypothetical prote
15	45	54.2	1255	2 T31065	diaphanous protein
16	44.5	53.6	635	2 F75477	hypothetical prote
17	44	53.0	405	2 A70845	probable penicilli
18	44	53.0	421	2 A56220	protein kinase [EC
19	44	53.0	430	2 JC7379	levansucrase (EC 2
20	44	53.0	569	2 A11347	hypothetical prote
21	43	51.8	312	2 T34688	probable lipoprote
22	43	51.8	376	2 S52137	hypothetical prote
23	43	51.8	402	2 S73773	MID2 protein - yea
24	43	51.8	812	2 S31521	dihydrolipoamide a
25	43	51.8	882	2 T23807	collagen COL1 - f
26	43	51.8	882	2 T23807	hypothetical prote
27	42.5	51.2	2761	2 T21064	hypothetical prote
28	42	50.6	130	2 S17961	BLT4 protein - bar
29	42	50.6	174	2 E84378	hypothetical prote

30	42	50.6	288	2 D75286	serine proteinase
31	42	50.6	393	2 JC5614	RNB6 protein - rat
32	42	50.6	510	2 H84887	probable pectinest
33	42	50.6	546	2 G72764	hypothetical prote
34	42	50.6	559	2 B55933	paxillin - chicken
35	42	50.6	586	2 T45945	laccase-like prote
36	42	50.6	630	2 S41314	hypothetical prote
37	42	50.6	671	2 A40692	signal recognition
38	42	50.6	839	2 T04859	extensin homolog F
39	42	50.6	974	1 A49714	protein-tyrosine k
40	42	50.6	994	2 I49276	C-mer tyrosine kin
41	42	50.6	1173	2 T31421	C-terminal domain-
42	42	50.6	1428	2 T08852	lustrin A - Califo
43	42	50.6	1568	2 T09074	semaphorin recepto
44	42	50.6	7463	2 T36248	CDA peptide synthe
45	41.5	50.0	491	2 F70699	probable pbpA prot

ALIGNMENTS

RESULT 1

S10571

mucin 1 precursor, secreted epithelial tumor antigen splice form - human  
N;Contains: mucin 1 secreted breast-cancer-associated splice form

C;Species: Homo sapiens (man)

C;Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text\_change 01-Dec-2000

C;Accession: S10571; JN0100; F56024; S09706; S10217

R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J  
Bur. J. Biochem. 189, 463-473, 1990

A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gener

A;Reference number: S10571; MUID:90276413; PMID:2351132

A;Accession: S10571

A;Molecule type: mRNA

A;Residues: 1-347 <WRE>

A;Cross-references: EMBL:X52228; NID:G36434; PIDN:CAA36477.1; PID:G36435

R;Tsarfaty, I.; Hareuveni, M.; Horev, J.; Zaretsky, J.; Weiss, M.; Jeltsch, J.M.; Garni

Gene 93, 313-318, 1990

A;Title: Isolation and characterization of an expressed hypervariable gene coding for a

A;Reference number: JN0100; MUID:91033045; PMID:1688329

A;Accession: JN0100

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-135, 'Q', 137-142, 'E', 144-163, 204-203, 'A', 210-347 <TSA>

A;Cross-references: GB:M35093; NID:G182252; PIDN:AAB59612.1; PID:G182253

R;Xing, P.X.; Tjandra, J.J.; Reynolds, K.; McLaughlin, P.J.; Purcell, D.F.J.; McKenzie,

J. Immunol. 142, 3503-3509, 1989

A;Title: Reactivity of anti-human milk fat globule antibodies with synthetic peptides.

A;Reference number: I56024; MUID:89235154; PMID:2715633

A;Accession: I56024

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 182-201 <RES>

A;Cross-references: GB:M26316; NID:G516622; PIDN:AAA36336.1; PID:G516623

R;Tandler, S.J.B.

Biochem. J. 267, 733-737, 1990

A;Title: Elements of secondary structure in a human epithelial mucin core peptide fragm

A;Reference number: S09706; MUID:90253387; PMID:2339983

A;Accession: S09706

A;Molecule type: protein

A;Residues: 182-201 <TEN>

C;Genetics:

A;Gene: GDB:MUCL; PUM

A;Cross-references: GDB:120705; OMIM:158340

A;Map position: 1q21-1q23

C;Keywords: alternative splicing; tandem repeat

F;1-23/Domain: signal sequence #status predicted <SIG>

F;24-347/Product: mucin 1, secreted epithelial tumor antigen splice form #status predict

Query Match 100.0%; Score 83; DB 2; Length 347;

Best Local Similarity 100.0%; Pred. No. 5.3e-05;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



QY 1 GVTSAPDTRPAPGSTA 16  
 Db 177 GVTSAPDTRPAPGSTA 192

RESULT 2  
 A35175  
 mucin 1 precursor, repetitive splice form A [validated] - human  
 N;Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin  
 ncreatic mucin; polymorphic epithelial mucin (PEM)  
 N;Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,  
 C;Species: Homo sapiens (man)  
 C;Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000  
 C;Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51  
 R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilken, J.  
 J. Biol. Chem. 265, 5573-5578, 1990  
 A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene end  
 A;Reference number: A35175; MUID:90202794; PMID:2318825  
 A;Accession: A35175

A;Molecule type: mRNA  
 A;Residues: 1-952,1033-1344 <LIG1>  
 A;Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124; GB:  
 A;Experimental source: splice form A  
 A;Note: GenBank entries HUMEPISIA1 and HUMEPISIA2 present only the amino-and carboxyl-en  
 A;Accession: B35175

A;Molecule type: mRNA  
 A;Residues: 1-19,29-952,1033-1344 <LIG2>  
 A;Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129; GB:  
 A;Experimental source: splice form B  
 A;Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl-en  
 R;Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel  
 J. Biol. Chem. 265, 15286-15293, 1990  
 A;Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli  
 A;Reference number: A35886; MUID:90368715; PMID:1697589  
 A;Accession: A35886

A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-19,29-992,1033-1344 <GEN>  
 A;Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870  
 A;Note: GenBank entry HUMUCAB includes one copy of the tandemly repeated sequence  
 R;Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.  
 J. Biol. Chem. 265, 15294-15299, 1990  
 A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.  
 A;Reference number: A35887; MUID:90368716; PMID:2394722  
 A;Accession: A35887

A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>  
 A;Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599  
 A;Note: GenBank entry HUMPANU contains four fewer copies of the tandemly repeated sequ  
 R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.  
 Eur. J. Biochem. 189, 463-473, 1990  
 A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera  
 A;Reference number: S10571; MUID:90276413; PMID:2351132  
 A;Accession: S10572

A;Molecule type: mRNA  
 A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>  
 A;Cross-references: EMBL:X52229; NID:g37053  
 R;Wreschner, D.H.  
 submitted to the EMBL Data Library, March 1990  
 A;Reference number: S40293  
 A;Accession: S40293

A;Molecule type: mRNA  
 A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WR2  
 A;Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054  
 R;Abe, M.; Siddiqui, J.; Kufe, D.  
 Biochem. Biophys. Res. Commun. 165, 644-649, 1989  
 A;Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated  
 A;Reference number: A36735; MUID:90088473; PMID:2597151  
 A;Accession: A36735

A;Molecule type: mRNA  
 A;Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>

A;Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543  
 R;Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu,  
 J. Biochem. 112, 609-615, 1992  
 A;Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agg.  
 A;Reference number: JX0235; MUID:93123189; PMID:1478919  
 A;Accession: PX0066

A;Molecule type: mRNA  
 A;Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>  
 A;Experimental source: gastric carcinoma cell  
 R;Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.  
 FEBS Lett. 356, 130-136, 1994  
 A;Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokerine  
 A;Reference number: S51026; MUID:95080414; PMID:7988707  
 A;Contents: annotation  
 A;Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region  
 C;Comment: This protein is length polymorphic. Individuals may have between 21 and 125  
 partial repeats. The repeat shown is defined by SmaI nuclease sites.  
 C;Comment: Serine and threonine residues in the tandem repeat domain are extensively g  
 C;Comment: For an alternative splice form without a tandem repeat domain, see PIR:S481.  
 C;Genetics:

A;Gene: GDB:MUC1; PUM  
 A;Cross-references: GDB:120705; OMIM:158340  
 A;Map position: 1q21-1q23  
 A;Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3  
 C;Superfamily: polymorphic epithelial mucin  
 C;Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorph  
 F;1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>  
 F;1-62/Region: mucin 1 amino-terminal non-repetitive  
 F;1-23/Domain: signal sequence #link PREA #status predicted <SIGA>  
 F;1-19,29-32/Domain: signal sequence #link PREA #status predicted <SIGB>  
 F;1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>  
 F;1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice fo  
 F;138-1017/Region: 20-residue repeats (GSTAPPAGGVTSAPDTRPAP)  
 F;1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive  
 F;1245-1272/Domain: transmembrane #status predicted <TRM>  
 F;1046,1064,1118,1144,1222/Binding site: carboxydrate (Asn) (covalent) #status predict  
 F;1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 83; DB 1; Length 1344;  
 Best Local Similarity 100.0%; Pred. No. 0.0002;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
 Db 146 GVTSAPDTRPAPGSTA 161

RESULT 3  
 A60533  
 tumor-associated antigen DF3 - human  
 C;Species: Homo sapiens (man)  
 C;Date: 19-Mar-1993 #sequence\_revision 07-May-1993 #text\_change 24-Nov-1999  
 C;Accession: A60533  
 R;Merlo, G.R.; Siddiqui, J.; Cropp, C.S.; Liscia, D.S.; Lidereau, R.; Callahan, R.; Ku  
 Cancer Res. 49, 6966-6971, 1989  
 A;Title: Frequent alteration of the DF3 tumor-associated antigen gene in primary human  
 A;Reference number: A60533; MUID:90058554; PMID:2582438  
 A;Accession: A60533  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-256 <MER>  
 C;Genetics:

A;Map position: 1q21-q24  
 C;Superfamily: proline-rich protein  
 C;Keywords: glycoprotein; tandem repeat

Query Match 83.1%; Score 69; DB 2; Length 256;  
 Best Local Similarity 86.7%; Pred. No. 0.0349;  
 Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VTSAPDTRPAPGSTA 16  
 Db 1 VTSAPESRPAPGSTA 15

```
RESULT 4
T50568
probable multi-domain regulatory protein [imported] - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T50568
R;Redenbach, M.; Kieser, H.M.; Denapaite, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopw
Mol. Microbiol. 21, 77-96, 1996
A;Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S
A;Reference number: Z20556; MUID:97000351; PMID:8843436
A;Accession: T50568
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1334 <RED>
A;Cross-references: EMBL:AL133220; PIDN:CAB61705.1
A;Experimental source: strain A3(2)
C;Genetics:
A;Note: SCC75A.05C

Query Match      68.7%; Score 57; DB 2; Length 1334;
Best Local Similarity 73.3%; Pred. No. 1.5;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAGST 15
Db 415 GTTAPGTAPAGST 429

RESULT 5
S73046
penicillin binding protein pbpC - Mycobacterium leprae
N;Alternate names: L308_f2_77 protein
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C;Accession: S73046
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid L308.
A;Reference number: S72590
A;Accession: S73046
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-182 <SMI>
A;Cross-references: EMBL:U00022; NID:G467164; PIDN:AAAL7345.1; PID:G467187
C;Genetics:
A;Gene: pbpC

Query Match      60.2%; Score 50; DB 2; Length 182;
Best Local Similarity 61.5%; Pred. No. 2.4;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAG 13
Db 88 GIVTAPDTPVPG 100

RESULT 6
D86995
probable D-alanyl-D-alanine carboxypeptidase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: D86995
R;Cole, S.T.; Eigelmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: D86995
A;Status: preliminary
```

```
A;Molecule type: DNA
A;Residues: 1-411 <STO>
A;Cross-references: GB:AL450380; NID:G13092840; PIDN:CAC30200.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0691

Query Match      60.2%; Score 50; DB 2; Length 411;
Best Local Similarity 61.5%; Pred. No. 5.3;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAPG 13
Db 88 GIVTAPDTPVPG 100

RESULT 7
T27876
hypothetical protein ZK470.2b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27876
R;Minx, P.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid ZK470.
A;Reference number: Z20433
A;Accession: T27876
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-78 <MIN>
A;Cross-references: EMBL:U39651; PIDN:AAA80396.1; CESP:ZK470.2b
A;Gene: CESP:ZK470.2b
A;Introns: 44/3

Query Match      59.0%; Score 49; DB 2; Length 78;
Best Local Similarity 83.3%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGS 14
Db 32 TSAPVITPAPGS 43

RESULT 8
T27875
hypothetical protein ZK470.2a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27875
R;Minx, P.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid ZK470.
A;Reference number: Z20433
A;Accession: T27875
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-175 <MIN>
A;Cross-references: EMBL:U39651; PIDN:AAA80395.1; CESP:ZK470.2a
C;Genetics:
A;Gene: CESP:ZK470.2a
A;Introns: 18/2; 55/3; 141/3

Query Match      59.0%; Score 49; DB 2; Length 175;
Best Local Similarity 83.3%; Pred. No. 3.3;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGS 14
Db 129 TSAPVITPAPGS 140

RESULT 9
T18264
```

cellulosome anchoring protein - Clostridium thermocellum  
C;Species: Clostridium thermocellum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T18264  
R;Fujino, T.; Beguin, P.; Aubert, J.P.  
J. Bacteriol. 175, 1891-1899, 1993  
A;Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulosome.  
e.  
A;Reference number: Z18847; MUID:93209931; PMID:8458832  
A;Accession: T18264  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-447 <FUJ>  
A;Cross-references: EMBL:X67506; NID:g296879; PID:g296883; PIDN:CAA47843.1  
C;Genetics:  
A;Gene: anca

Query Match 59.0%; Score 49; DB 2; Length 447;  
Best Local Similarity 69.2%; Pred. No. 8.1;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 SAPDTRPAPGSTA 16  
|||: |||||  
Db 176 SAPRATPTPGSTA 188

RESULT 10  
G70951  
probable ATP-dependent DNA helicase - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: G70951  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: G70951  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1101 <COL>  
A;Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16666.1; PID:e124879  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: RV3201c

Query Match 59.0%; Score 49; DB 2; Length 1101;  
Best Local Similarity 56.2%; Pred. No. 20;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAPGSTA 16  
||: |||||  
Db 1073 GIVVPDELPAAGELA 1088

RESULT 11  
E75277  
ABC transporter, ATP-binding protein, Msba family - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C;Accession: E75277  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, X.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896; PMID:10567266  
A;Accession: E75277  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-633 <WHI>

A;Cross-references: GB:AE002071; GB:AE000513; NID:g6460218; PIDN:AAF11948.1; PID:g6460;  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR2404  
A;Map position: 1  
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology e.  
Query Match 56.6%; Score 47; DB 2; Length 633;  
Best Local Similarity 75.0%; Pred. No. 23;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAP 12  
|||: |||||  
Db 366 GVTLPDPRPLP 377

RESULT 12  
F72653  
hypothetical protein APE0658 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C;Accession: F72653  
R;Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: F72653  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-108 <KAW>  
A;Cross-references: DDBJ:AP000060; NID:g5104183; PIDN:BAA79630.1; PID:dl043416; PID:g5104183  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE0658

Query Match 55.4%; Score 46; DB 2; Length 108;  
Best Local Similarity 69.2%; Pred. No. 5.7;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAPG 13  
|||: |||||  
Db 79 GVGSHPGSRPAPG 91

RESULT 13  
T18853  
probable transcription factor - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
C;Accession: T18853; T23730  
R;Berks, M.  
submitted to the EMBL Data Library, April 1996  
A;Reference number: Z19030  
A;Accession: T18853  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-399 <WIL>  
A;Cross-references: EMBL:Z71258; PIDN:CAA95787.1; GSPDB:GN00019; CESP:M05B5.5  
A;Experimental source: clone C01H6  
R;Gardner, A.  
submitted to the EMBL Data Library, April 1996  
A;Reference number: Z19790  
A;Accession: T23730  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-399 <W12>  
A;Cross-references: EMBL:Z71265; PIDN:CAA95837.1; GSPDB:GN00019; CESP:M05B5.5  
A;Experimental source: clone M05B5  
C;Genetics:  
A;Gene: CESP:M05B5.5  
A;Map position: 1  
A;Introns: 49/1; 211/3; 270/2; 312/2; 359/3



A;Description: Mutations in aurora prevent centrosome separation leading to the formation of a multipolar spindle

A;Accession: S51673

A;Reference: S51673

A;Molecule type: mRNA

A;Residues: 1-421 <GL2>

A;Cross-references: EMBL:X83465; NID:G603534; PIDN:CAA58468.1; PID:G603535

A;Accession: S51674

A;Molecule type: DNA

A;Residues: 1-421 <GLO>

A;Cross-references: EMBL:X83466; NID:G603536; PIDN:CAA58469.1; PID:G603537

C;Genetics:

A;Gene: aur

A;Cross-references: FlyBase:FBgn0000147

A;Map position: 87A7-A9

A;Introns: 38/3; 149/1

A;Note: aurora

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

C;Keywords: ATP; phosphotransferase

F;162-416/Domain: protein kinase homology <KIN>

F;170-178/Region: protein kinase ATP-binding motif

Query Match 53.0%; Score 44; DB 2; Length 421;  
Best Local Similarity 63.6%; Pred. No. 43;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 APDTRPAPGST 15  
||||:|||||

Db 45 APDSKPLPGSS 55

RESULT 19

JC7379

levansucrase [EC 2.4.1.10] - Acetobacter xylinum, Acetobacter aceti subsp. xylinum

C;Species: Acetobacter xylinum, Acetobacter aceti subsp. xylinum

C;Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 31-Dec-2000

C;Accession: JC7379

R;Tajima, K.; Tanio, T.; Kobayashi, Y.; Kohno, H.; Fujiwara, M.; Shiba, T.; Ezata, T.; M  
DNA Res. 7, 237-242, 2000

A;Title: Cloning and sequencing of the levansucrase gene from Acetobacter xylinum NCI 10

A;Reference number: JC7379

A;Accession: JC7379

A;Molecule type: DNA

A;Residues: 1-430 <TAJ>

A;Cross-references: DDBJ:AB034152

A;Experimental source: strain NCI 1005

C;Comment: This enzyme, having high polymerase activity, releases glucose and produces 1

C;Genetics:

A;Gene: lsx4

C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 53.0%; Score 44; DB 2; Length 430;  
Best Local Similarity 71.4%; Pred. No. 44;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 VTSAPDTRPAPGST 15  
|||:|||||

Db 371 VQSFIDTLAPGST 384

RESULT 20

AI1347

hypothetical protein Imo2185 [imported] - Listeria monocytogenes (strain EGD-e)

C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AI1347

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AI1347

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-569 <GLA>

A;Cross-references: GB:NC\_003210; PIDN:CAD00263.1; PID:G16411655; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lmo2185

Query Match 53.0%; Score 44; DB 2; Length 569;  
Best Local Similarity 53.3%; Pred. No. 58;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGST 15  
|||:|||||

Db 323 GETNFWTKPDPGTT 337

RESULT 21

T34688

probable lipoprotein - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C;Accession: T34688

R;Harris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1998

A;Reference number: Z21553

A;Accession: T34688

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-289 <HAR>

A;Cross-references: EMBL:AL023517; PIDN:CAA18984.1; GSPDB:GN00070; SCOEDB:SC1B5.10C

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SC1B5.10C

Query Match 51.8%; Score 43; DB 2; Length 289;  
Best Local Similarity 61.5%; Pred. No. 42;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 SAPDTRPAPGSTA 16  
|||||:|:|

Db 35 SAPDFSPSPSSA 47

RESULT 22

AB2063

hypothetical protein all2056 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AB2063

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch  
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,  
DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AB2063

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-312 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA073755.1; PID:G17131147; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all2056

C;Superfamily: tropinesterase

Query Match 51.8%; Score 43; DB 2; Length 312;  
Best Local Similarity 58.3%; Pred. No. 45;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 VTSAPDTRPAPG 13  
:|:|:|



Db 3 TTAPSTTPIPG 14

RESULT 23

MID2 protein - yeast (*Saccharomyces cerevisiae*)  
N;Alternate names: KAI1 protein; protein L8543.19; protein YLR332w; serine-rich transmembrane protein  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-2000  
C;Accession: S52137; S52073; S53408; S35261  
R;Ono, T.; Suzuki, T.; Anraku, Y.; Iida, H.  
A;Submitted to the EMBL Data Library, April 1994  
A;Description: Cloning and characterization of the MID2 gene required for the mating process in *S. cerevisiae*  
A;Reference number: S52137  
A;Accession: S52137  
A;Molecule type: DNA  
A;Residues: 1-376 <ONO>  
A;Cross-references: EMBL:D29945; NID:g479040; PID:dl006784; PID:g479041  
R;Takeuchi, J.; Okada, M.; Toh-e, A.; Kikuchi, Y.  
Biochim. Biophys. Acta 1260, 94-96, 1995  
A;Title: The SMS1 gene encoding a serine-rich transmembrane protein suppresses the temperature sensitivity of a mating defect in *S. cerevisiae*  
A;Reference number: S52073; MUID:95092800; PMID:7999801  
A;Accession: S52073  
A;Molecule type: DNA  
A;Residues: 1-376 <TAK>  
A;Cross-references: EMBL:D29964; NID:g479042; PIDN:BAA06230.1; PID:g676885  
R;Du, Z.

submitted to the EMBL Data Library, February 1995  
A;Description: The sequence of *S. cerevisiae* cosmid 8543.  
A;Reference number: S53390  
A;Accession: S53408  
A;Molecule type: DNA  
A;Residues: 1-376 <DUZ>  
A;Cross-references: EMBL:U20618; NID:G2258165; PID:G662143; MIPS:YLR332W  
R;Daniel, J.  
Mol. Genet. 240, 245-257, 1993  
A;Title: Potentially rapid walking in cellular regulatory networks using the gene-gene i  
A;Reference number: S35260; MUID:93360904; PMID:8355657  
A;Accession: S35261  
A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA  
A;Residues: 60-302,'VV' <DAN>  
A;Note: the sequence from Fig. 7B is inconsistent with that from Fig. 7A in lacking 138-139  
C;Genetics:  
A;Gene: SGD:MID2; SMS1; KAI1  
A;Cross-references: SGD:S0004324; MIPS:YLR332W  
A;Map position: 12R  
C;Keywords: transmembrane protein  
F;13-29/Domain: transmembrane #status predicted <TM1>  
F;229-245/Domain: transmembrane #status predicted <TM2>

Query Match 51.8%; Score 43; DB 2; Length 376;  
Best Local Similarity 53.3%; Pred. No. 54;  
Matches 8: Conservative 2; Mismatches 5; Indels 0; Gaps 0

2 VTSAPDTRPAPGSTA 16

db 161 ITSAPSTSSTPSTA 175

RESULT 24

Dihydrolipoamide acetyltransferase component E2 - Mycoplasma pneumoniae (strain ATCC 29321)  
 N/A; Alternate names: hypothetical protein Fli\_orf402  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342  
 C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
 C:Accession: S73773  
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkle, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
 A:Reference number: S73327; MUID:97105885; PMID:8948633  
 A:Accession: S73773

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-402 <HIM>

A;Cross-references: EMBL:AE000044; GB:U00089; NID:g1674130; PIDN:AAB96095.1; PID:g16741  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
C;Genetics:

A:Genetic code: SGC3  
C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology  
F:4-77/Domain: lipoyl/biotin-binding homology <LPB>

Query Match 51.8%; Score 43; DB 2; Length 402;  
Best Local Similarity 64.3%; Pred.No. 58;  
Matches 9; Conservative 0; Mismatches 5; Indels

Qy 3 TSAPDTRPAPGSTA 16

Db 154 TSAPTTPAPASA 167

RESULT 25

collagen COLF1 - freshwater sponge (*Ephydatia muelleri*)  
C;Species: *Ephydatia muelleri*

C:\Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 15-Sep-2003  
C:\Accession: S31521

F;Exposito, J.Y.; van der Rest, M.; Garrone, R.  
submitted to the EM3L Data Library, December 1992  
A;Description: The complete intron/exon structure of E. muelleri collagen gene suggests  
A;Reference number: S31521

A;Accession: S31521

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-812 <EXP>  
A;Cross-references: EMBL:X69818; NID:g429036; PIDN:CAA49472.1; PFD:g9300

Query Match 51.8%; Score 43; DB 2; Length 812;  
Best Local Similarity 60.0%; Pred. No. 1.1e-02;  
Matches 9; Conservative 0; Mismatches 6; Indels

Qy 1 GVTSAPDTRPAPGST 15

Db 441 GVTGAPGKEGRPGST 455

## RESULT 26

T23807  
hypothetical protein M28.4 - *Caenorhabditis elegans*

```
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23807
```

R; Gardner, A.

submitted to the EMEL Data Library, June 1995

A;Reference number: Z19803

**A;Accession:** T23807

A;Status: preliminary; translated from GB/EMBL/DD8J

A;Molecule type: DNA.

A;KELQUES: I-882 <WILL>

A:Cross-References: EMBL:Z49911; PIDN:CAA90I26.1; GSPDB:GN00020; CESP:M28.1  
A:Experimental source: clone M28

C; Genetics:

A; Gene: CBSP:M28.4

A;Map position: 2

A; Introns: 67/3; 130/3; 183/1; 236/3; 289/3; 317/3; 372/1; 452/2; 512/3; 573/3; 600/3;

### Query Match

Query Match 51.8%; Score 43; DB 2; Length 882;  
Best Local Similarity 60.0%;  
Pred. No. 1.2e+02;  
Matches 9; Conservative  
2; Mismatches 4; Indels

2cy 1 GVTSA PDTRPAPGST 15

815 GSTAAPVSTDAPGST 829



```
RESULT 27
T21064
hypothetical protein F53B7.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21064; T22550
R:McMurray, A.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19366
A:Accession: T21064
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2761 <WIL>
A:Cross-references: EMBL:Z72507; PIDN:CAA96634.1; GSPDB:GN00023; CESP:F53B7.5
A:Experimental source: clone F17C11
R:McMurray, A.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19579
A:Accession: T22550
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2761 <W12>
A:Cross-references: EMBL:Z72510; PIDN:CAA96654.1; GSPDB:GN00023; CESP:F53B7.5
A:Experimental source: clone F53B7
C:Genetics:
A:Gene: CESP:F53B7.5
A:Map position: 5
A:Introns: 498/1; 543/3; 656/3; 904/1; 1018/1; 1148/1; 1245/1; 1433/3; 1616/1; 1661/3; 1
Query Match 51.2%; Score 42.5; DB 2; Length 2761;
Best Local Similarity 61.1%; Pred. No. 4.5e+02;
Matches 11; Conservative 0; Mismatches 4; Indels 3; Gaps 1;
QY 1 GVTS--APDTRPAPGST 15
||| |||||
DB 2537 GATSTSPAPITTPAPSST 2554

RESULT 28
S17961
BLT4 protein - barley
C:Species: Hordeum vulgare (barley)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C:Accession: S17961
R:Dunn, M.A.; Hughes, M.A.; Zhang, L.; Pearce, R.S.; Quigley, A.S.; Jack, P.L.
Mol. Gen. Genet. 229, 389-394, 1991
A:Title: Nucleotide sequence and molecular analysis of the low temperature induced cereal
A:Reference number: S17961; MUID:92049237; PMID:1944226
A:Accession: S17961
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-130 <DUN>
A:Cross-references: GB:X56547; NID:g18856; PIDN:CAA39887.1; PID:g18857
C:Superfamily: phospholipid transfer protein
Query Match 50.6%; Score 42; DB 2; Length 130;
Best Local Similarity 62.5%; Pred. No. 27;
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GVTSAPDTRPAPGSTA 16
| ||| |||
DB 95 GAASAPTRSAPVSTA 110

RESULT 29
E84378
hypothetical protein Vng2279h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84378
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
```

```
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: E84378
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-174 <STO>
A:Cross-references: GB:AE004437; NID:g10581690; PIDN:AAG20393.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG2279H
```

```
Query Match 50.6%; Score 42; DB 2; Length 174;
Best Local Similarity 57.1%; Pred. No. 36;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 VTSAPDTRPAPGST 15
||: ||| | | | : |
DB 117 VTTAADTEPGPPTT 130
```

## RESULT 30

```
D75286
serine proteinase truncated homolog DR2322 [imported] - Deinococcus radiodurans (strain)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: D75286
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: D75286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-288 <WHI>
A:Cross-references: GB:AE002064; GB:AE000513; NID:g6460134; PIDN:AAF11870.1; PID:g6460134;
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2322
A:Map position: 1
```

```
Query Match 50.6%; Score 42; DB 2; Length 288;
Best Local Similarity 61.5%; Pred. No. 59;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 3 TSAPCTRPAPGST 15
| ||| ||| | | |
DB 181 TPAPTTPAPGTT 193
```

## RESULT 31

```
JC5614
RNB6 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 23-Sep-1997 #sequence_revision 17-Oct-1997 #text_change 05-Nov-1999
C:Accession: JC5614
R:Ohta, S.; Mineta, T.; Kimoto, M.; Tabuchi, K.
Biochem. Biophys. Res. Commun. 237, 307-312, 1997
A:Title: Differential display cloning of a novel rat cDNA (RNB6) that shows high expres
A:Reference number: JC5614; MUID:97415794; PMID:9268706
A:Accession: JC5614
A:Molecule type: mRNA
A:Residues: 1-393 <OHT>
A:Cross-references: GB:U70211; NID:g2058461; PIDN:AAC53322.1; PID:g2058462
A:Experimental source: brain
C:Comment: This protein belongs to Ena/VASP family member, and is involved in the devei
ion.
```

```
Query Match 50.6%; Score 42; DB 2; Length 393;
Best Local Similarity 57.1%; Pred. No. 80;
```

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TSAPDTRPAPGSTA 16  
| | | | |  
Db 296 TEDPSTSPSGSRA 309  
| | | | |

RESULT 32  
H84887  
probable pectinesterase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: H84887  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: H84887  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-510 <STO>  
A:Cross-references: GB:AE002093; NID:g2583131; PIDN:AAB82640.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g45220  
A:Map position: 2  
C:Superfamily: pectinesterase

Query Match 50.6%; Score 42; DB 2; Length 510;  
Best Local Similarity 64.3%; Pred. No. 1e+02;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 VTSAPDTRPAPGST 15  
| | | | |  
Db 398 VTAASDLRFVLGST 411  
| | | | |

RESULT 33  
G72764  
hypothetical protein APE0107 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: G72764  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: G72764  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-546 <KAW>  
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79017.1; PID:g5103496  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0107  
C:Superfamily: conserved hypothetical protein MJ1429

Query Match 50.6%; Score 42; DB 2; Length 546;  
Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GVTSAPDTRPAPGST 15  
| | | | |  
Db 111 GEQSAKIPPPQPGST 125  
| | | | |

RESULT 34  
B55933  
paxillin - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 21-Jul-2000

C:Accession: B55933  
R:Salgia, R.; Li, J.L.; Lo, S.H.; Brunkhorst, B.; Kansas, G.S.; Sobhany, E.S.; Sun, Y.;  
J. Biol. Chem. 270, 5039-5047, 1995  
A:Title: Molecular cloning of human paxillin, a focal adhesion protein phosphorylated b  
A:Reference number: A55933; MUID:95197488; PMID:7534286  
A:Accession: B55933  
A:Molecule type: mRNA  
A:Residues: 1-559 <SAL>  
A:Cross-references: GB:U14589; NID:g704349; PIDN:AAC59665.1; PID:g704350  
C:Superfamily: LIM metal-binding repeat homology  
C:Keywords: Cell adhesion; cytoskeleton; phosphoprotein; zinc finger  
P:46-55/Region: proline-rich  
F:326-376/Domain: LIM metal-binding repeat homology <LIM1>  
F:385-435/Domain: LIM metal-binding repeat homology <LIM2>  
F:444-494/Domain: LIM metal-binding repeat homology <LIM3>  
F:503-553/Domain: LIM metal-binding repeat homology <LIM4>  
F:31,118/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 50.6%; Score 42; DB 2; Length 559;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GVTSAPDTRPAPGS 14  
| | | | |  
Db 287 GSSSPSTTPKPGS 300  
| | | | |

RESULT 35  
T45945  
laccase-like protein - Arabidopsis thaliana  
N:Alternate names: protein F7J8.30  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T45945  
R:Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Len  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23018  
A:Accession: T45945  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-586 <EEV>  
A:Cross-references: EMBL:AL137189  
A:Experimental source: cultivar Columbia; BAC clone F7J8  
C:Genetics:  
A:Map position: 5  
A:Introns: 33/3; 84/2; 166/1; 203/1; 526/1  
A:Note: F7J8.30  
C:Superfamily: laccase

Query Match 50.6%; Score 42; DB 2; Length 586;  
Best Local Similarity 53.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVTSAPDTRPAPG 13  
| | | | |  
Db 288 GVPASPDTRKTRG 300  
| | | | |

RESULT 36  
S41314  
hypothetical protein P2 - cucurbit aphid-borne yellows virus  
C:Species: cucurbit aphid-borne yellows virus  
C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 20-Sep-1999  
C:Accession: S41314  
R:Guilley, H.; Wipf-Scheibel, C.; Richards, K.; Jecocq, H.; Jonard, G.  
submitted to the EMBL Data Library, December 1993  
A:Description: Nucleotide sequence of cucurbit aphid-borne yellows virus.  
A:Reference number: S41313  
A:Accession: S41314  
A:Molecule type: genomic RNA  
A:Residues: 1-630 <GUI>  
A:Cross-references: EMBL:X76931; NID:g441268; PIDN:CAA54250.1; PID:g441270  
C:Superfamily: potato leaf roll virus 70K protein

```

Query Match          50.6%; Score 42; DB 2; Length 630;
Best Local Similarity 57.1%; Pred. No. 1.3e+02;
Matches      8; Conservative    2; Mismatches   4; Indels   0; Gaps   0;

QY       3 TSAPDTRPAGSTA 16
         |:||| ||||| :
Db        458 TTAPDATPAVGRTS 471

RESULT 37
A40692
signal recognition particle 72K chain - dog
C/Species: Canis lupus familiaris (dog)
C/Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C/Accession: A40692; S32167
R/Lutcke, H.; Prehn, S.; Ashford, A.J.; Remus, M.; Frank, R.; Dobberstein, B.
J. Cell Biol. 121, 977-985, 1993
A/Title: Assembly of the 68- and 72-kD proteins of signal recognition particle with 7S
A/Reference number: A40692; PMID:8388879
A/Accession: A40692
A/Status: preliminary
A/Molecule type: mRNA; protein
A/Residues: 1-671 <LUT>
A/Cross-references: EMBL:X67813; NID:g297767; PIDN:CAA48014.1; PID:g297768
A/Note: sequence extracted from NCBI backbone (NCBIP:132901)

Query Match          50.6%; Score 42; DB 2; Length 671;
Best Local Similarity 61.5%; Pred. No. 1.3e+02;
Matches      8; Conservative    1; Mismatches   4; Indels   0; Gaps   0;

QY       4 SAPDTRPREGSTA 16
         |:||| ||||| :
Db        620 SSPPTSPREGSAA 632

RESULT 38
T04859
extensin homolog F28A21.80 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C/Accession: T04859
R/Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer,
submitted to the Protein Sequence Database, February 1999
A/Reference number: Z15387
A/Accession: T04859
A/Molecule type: DNA
A/Residues: 1-839 <BEV>
A/Cross-references: EMBL:AL035526
A/Experimental source: cultivar Columbia; BAC clone F28A21
C/Genetics:
A/Map position: 4
A/Introns: 623/3
A/Note: F28A21.80

Query Match          50.6%; Score 42; DB 2; Length 839;
Best Local Similarity 61.5%; Pred. No. 1.7e+02;
Matches      8; Conservative    1; Mismatches   4; Indels   0; Gaps   0;

QY       2 VTSAPDTRPAPGS 14
         |:||| ||||| :
Db        455 VPSPSTTPSPGS 467

RESULT 39
A49714
protein-tyrosine kinase (EC 2.7.1.112) c-eyk precursor - chicken
C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 04-Feb-2000
C/Accession: A49714
R/Jia, R.; Hanafusa, H.
J. Biol. Chem. 269, 1839-1844, 1994
A/Title: The proto-oncogene of v-eyk (v-ryk) is a novel receptor-type protein tyrosine k

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:22:19 ; Search time 8.61539 Seconds  
(without alignments)  
96.702 Million cell updates/sec

Title: US-10-070-566-6

Perfect score: 83

Sequence: 1 GVTSAPDTRPAPGSTA 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query %	Description
1	83	100.0	1255	1 MUC1_HUMAN
2	69	83.1	475	1 MUC1_HYLLA
3	52	62.7	333	1 NK32_HUMAN
4	49	59.0	447	1 ANCA_CLOTM
5	47	56.6	564	1 ZYX_MOUSE
6	46	55.4	145	1 CUB8_HUMAN
7	45	54.2	1255	1 DIA1_MOUSE
8	44	53.0	333	1 NK32_MOUSE
9	44	53.0	351	1 FXI1_HUMAN
10	44	53.0	990	1 AMPM_MANSE
11	43	51.8	303	1 PCPR_SPHCR
12	43	51.8	309	1 CIWG_HUMAN
13	43	51.8	376	1 MID2_YEAST
14	43	51.8	402	1 ODP2_MYCPN
15	42	50.6	130	1 BLT4_HORVU
16	42	50.6	393	1 EVL_RAT
17	42	50.6	559	1 PAXI_CHICK
18	42	50.6	670	1 SR72_CANFA
19	42	50.6	670	1 SR72_HUMAN
20	42	50.6	994	1 MERK_MOUSE
21	42	50.6	994	1 MERK_RAT
22	41	49.4	225	1 UNG1_STRCO
23	41	49.4	476	1 TRZA_RHOCO
24	41	49.4	559	1 ECM1_MOUSE
25	41	49.4	747	1 GUND_CELFI
26	41	49.4	1150	1 APMU_PIG
27	41	49.4	1216	1 PIB1_MOUSE
28	41	49.4	1216	1 PIB1_RAT
29	41	49.4	1723	1 AIM1_HUMAN
30	41	49.4	4391	1 PGBM_HUMAN
31	41	49.4	5038	1 PCLO_MOUSE
32	40.5	48.8	3301	1 CLR3_MOUSE
33	40.5	48.8	3313	1 CLR3_RAT

34	40	48.2	196	1 BCE1_ARATH	Q07488 arabidopsis
35	40	48.2	227	1 UNG_MYCLE	Q9cbs3 mycobacteri
36	40	48.2	302	1 GLTI_ECOLI	P37902 escherichia
37	40	48.2	336	1 LIPB_DEIRA	Q9rwa5 deinococcus
38	40	48.2	364	1 FMLR_MOUSE	P33766 mus musculus
39	40	48.2	385	1 MS1A_HUMAN	O14940 homo sapien
40	40	48.2	398	1 ODO2_AZOVI	P20708 azotobacter
41	40	48.2	419	1 KAS2_STRCN	Q02579 streptomyce
42	40	48.2	449	1 GSNB_DROME	P09083 drosophila
43	40	48.2	509	1 SHS1_RAT	P97710 r protein-t
44	40	48.2	629	1 SYR_PYRFU	Q8ul49 pyrococcus
45	40	48.2	636	1 MS1B_HUMAN	Q9nzb8 homo sapien

# ALIGNMENTS

RESULT 1					
MUC1_HUMAN					
ID	MUC1_HUMAN	STANDARD;	PRT;	1255 AA.	
AC	P15941;	P13931;	P15942;	P17626;	Q14128; Q14876; Q16437; Q16442;
AC	Q16615;	Q9BXA4;	Q9UE75;	Q9UE76;	Q9Y4J2;
DT	01-JAN-1990	(Rel. 13, Created)			
DT	01-APR-1990	(Rel. 14, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Mucin 1 precursor (MUC-1) {Polymorphic epithelial mucin} (PEM) (PEMT)				
DE	{Episialin} (Tumor-associated mucin) (Carcinoma-associated mucin)				
DE	{Tumor-associated epithelial membrane antigen} (EMA) (H23AG) (Peanut-				
DE	reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen				
DE	DF3) {CD227 antigen}.				
GN	MUC1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RC	TISSUE=Pancreas;				
RX	MEDLINE=90368716; PubMed=2394722;				
RA	Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;				
RT	"Cloning and sequencing of a human pancreatic tumor mucin cDNA.";				
RL	J. Biol. Chem. 265:15294-15299(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).				
RX	MEDLINE=90202794; PubMed=2318825;				
RA	Ligtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilken J.;				
RT	"Episialin, a carcinoma-associated mucin, is generated by a				
RT	polymorphic gene encoding splice variants with alternative amino				
RT	termini.";				
RL	J. Biol. Chem. 265:5573-5578(1990).				
RN	[3]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RC	TISSUE=Breast carcinoma;				
RX	MEDLINE=90368715; PubMed=1697589;				
RA	Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,				
RA	Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;				
RT	"Molecular cloning and expression of human tumor-associated				
RT	polymorphic epithelial mucin.";				
RL	J. Biol. Chem. 265:15286-15293(1990).				
RN	[4]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RX	MEDLINE=91097524; PubMed=2268309;				
RA	Lancaster C.A., Peat N., Duhig T., Wilson D.;				
RA	Taylor-Papadimitriou J., Gendler S.J.;				
RT	"Structure and expression of the human polymorphic epithelial mucin				
RT	gene: an expressed VNTR unit.";				
RL	Biochem. Biophys. Res. Commun. 173:1019-1029(1990).				
RN	[5]				
RP	SEQUENCE FROM N.A. (ISOFORM 5).				
RC	TISSUE=Breast carcinoma;				
RX	MEDLINE=90276413; PubMed=2351132;				
RA	Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,				
RA	Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;				

RT "Human epithelial tumor antigen cDNA sequences. Differential splicing  
RT may generate multiple protein forms.";  
RL Eur. J. Biochem. 189:463-473(1990).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=90276414; PubMed=2112460;  
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,  
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;  
RT "A transcribed gene, containing a variable number of tandem repeats,  
RT codes for a human epithelial tumor antigen. cDNA cloning, expression  
RT of the transfected gene and over-expression in breast cancer  
RT tissue.";  
RL Eur. J. Biochem. 189:475-486(1990).  
RN [7]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=91033045; PubMed=1688329;  
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,  
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;  
RT "Isolation and characterization of an expressed hypervariable gene  
RT coding for a breast-cancer-associated antigen.";  
RL Gene 93:313-318(1990).  
RN [8]  
RP SEQUENCE FROM N.A. (ISOFORM 7).  
RX MEDLINE=95010060; PubMed=7925397;  
RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,  
RA Keydar I., Hilkens J., Wreschner D.H.;  
RT "Characterization and molecular cloning of a novel MUC1 protein,  
RT devoid of tandem repeats, expressed in human breast cancer tissue.";  
RL Eur. J. Biochem. 224:787-795(1994).  
RN [9]  
RP SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).  
RX MEDLINE=97355747; PubMed=9212228;  
RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,  
RA Finstad C.J.;  
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial  
RT cancer cell lines and demonstration of a new short variant form  
RT (MUC-1/Z).";  
RL Int. J. Cancer 72:87-94(1997).  
RN [10]  
RP SEQUENCE FROM N.A. (ISOFORM 7).  
RX Zhang L.X., Li C.H.;  
RT "Molecular cloning of an isoform of MUC1, MUC1/Y.";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [11]  
RP SEQUENCE FROM N.A. (ISOFORM 9).  
RC TISSUE=Epithelial cancer;  
RA Zhang L.X., Li C.H., Sun L.Y., Yue W.;  
RT "Cloning of a new potential secreted short variant form of MUC1 mucin  
RT in epithelial cancer cell line.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [12]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=88330762; PubMed=3417635;  
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,  
RA Burchell J.;  
RT "A highly immunogenic region of a human polymorphic epithelial mucin  
RT expressed by carcinomas is made up of tandem repeats.";  
RL J. Biol. Chem. 263:12820-12823(1988).  
RN [13]  
RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM 2).  
RX MEDLINE=90088473; PubMed=2597151;  
RA Abe M., Siddiqui J., Kufe D.;  
RT "Sequence analysis of the 5' region of the human DF3 breast  
RT carcinoma-associated antigen gene.";  
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).  
RN [14]  
RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).  
RC TISSUE=Thyroid;  
RX MEDLINE=96183746; PubMed=8608966;  
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;  
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse  
RT transcriptase polymerase chain reaction of the MUC1 gene.";

RL Int. J. Cancer 66:55-59(1996).  
RN [15]  
RP SEQUENCE OF 1-89 FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=96181716; PubMed=8604237;  
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,  
RA Lee L.N., Luh K.T., Wu C.W.;  
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and  
RT tissues.";  
RL Oncology 53:118-126(1996).  
RN [16]  
RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).  
RC TISSUE=Breast carcinoma;  
RA Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;  
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
RN [17]  
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.  
RX MEDLINE=97460054; PubMed=9312074;  
RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,  
RA Hanisch F.-G.;  
RT "Localization of O-glycosylation sites on glycopeptide fragments from  
RT lactation-associated MUC1. All putative sites within the tandem  
RT repeat are glycosylation targets in vivo.";  
RL J. Biol. Chem. 272:24780-24793(1997).  
RN [18]  
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.  
RX MEDLINE=99303572; PubMed=10373415;  
RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,  
RA Hanisch F.-G.;  
RT "High density O-glycosylation on tandem repeat peptide from secretory  
RT MUC1 of T47D breast cancer cells.";  
RL J. Biol. Chem. 274:18165-18172(1999).  
RN [19]  
RP POLYMORPHISM WITHIN THE REPEAT.  
RX MEDLINE=21359366; PubMed=11350974;  
RA Engelmann K., Baldus S.B., Hanisch F.-G.;  
RT "Identification and topology of variant sequences within individual  
RT repeat domains of the human epithelial tumor mucin MUC1.";  
RL J. Biol. Chem. 276:27764-27769(2001).  
RN [20]  
RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.  
RX MEDLINE=99211485; PubMed=10197628;  
RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,  
RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,  
RA Wreschner D.H.;  
RT "The breast cancer-associated MUC1 gene generates both a receptor and  
RT its cognate binding protein.";  
RL Cancer Res. 59:1552-1561(1999).  
RN [21]  
RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.  
RX MEDLINE=21240104; PubMed=11341784;  
RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,  
RA Harris A.;  
RT "Identification of MUC1 proteolytic cleavage sites in vivo.";  
RL Biochem. Biophys. Res. Commun. 283:715-720(2001).  
RN [22]  
RP CHARACTERIZATION.  
RX MEDLINE=21836452; PubMed=11847293;  
RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,  
RA Ziv R., Okun I., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,  
RA Stacey M., Lin H.-H., Gordon S.;  
RT "Generation of ligand-receptor alliances by 'SEA' module-mediated  
RT cleavage of membrane-associated mucin proteins.";  
RL Protein Sci. 11:698-706(2002).  
RN [23]  
RP PHOSPHORYLATION.  
RX MEDLINE=95080414; PubMed=7988707;  
RA Zrihan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.;  
RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.  
RT Cytokine receptor-like molecules.";  
RL FEBS Lett. 356:130-136(1994).  
CC - - FUNCTION: May play a role in adhesive functions and in cell-cell  
CC interactions, metastasis and signaling. May provide a protective



Query Match 100.0%; Score 83; DB 1; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 0.00014;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAGSTA 16  
| | | | | | | | | | | | | | | | | |  
Db 137 GVTSAPDTRPAGSTA 152

RESULT 2  
MUC1\_HYLLA  
ID MUC1\_HYLLA STANDARD; PRT; 475 AA.  
AC Q29435;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Mucin 1 precursor (MUC-1).  
GN MUC1.  
OS Hylobates lar (Common gibbon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hylobatidae; Hylobates.  
OX NCBI\_TaxID=9580;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96351712; PubMed=8747930;  
RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;  
RT "Analysis of mammalian MUC1 genes reveals potential functionally  
important domains";  
RL Mamm. Genome 6:885-888(1995).  
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
CYTOSKELETON (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.  
CC -!- SIMILARITY: Contains 1 SEA domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L41589; AAA69965.1; -;  
EMBL; L41625; AAA69918.1; -;  
EMBL; L41624; AAA69918.1; JOINED.  
InterPro: IPR000082; SEA\_domain.  
Pfam; PF01390; SEA; 1.  
SMART; SM00200; SEA; 1.  
PROSITE; PS50024; SEA; 1.  
KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  
Repeat.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 475 MUCIN 1.  
FT DOMAIN 24 380 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 381 401 POTENTIAL.  
FT DOMAIN 402 475 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 102 121 1.  
FT REPEAT 122 141 2.  
FT REPEAT 142 161 3.  
FT REPEAT 162 181 4.  
FT DOMAIN 254 371 SEA.  
SQ SEQUENCE 475 AA; 49371 MW; D7A699D6D68C6522 CRC64;

Query Match 83.1%; Score 69; DB 1; Length 475;  
Best Local Similarity 87.5%; Pred. No. 0.006;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAGSTA 16  
| | | | | | | | | | | | | | | | | |  
Db 137 GVTSAPDTRPAGSTA 152

RESULT 3  
NK32\_HUMAN  
ID NK32\_HUMAN STANDARD; PRT; 333 AA.  
AC P78367;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Homeobox protein Nkx-3.2 (Bagpipe homeobox protein homolog 1).  
GN BAPX1 OR NKX3B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98086223; PubMed=9426254;  
RA Tribioli C., Lufkin T.;  
RT "Molecular cloning, chromosomal mapping and developmental expression  
of BAPX1, a novel human homeobox-containing gene homologous to  
Drosophila bagpipe";  
RL Gene 203:225-233(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98008936; PubMed=93444671;  
RA Yoshiura K.I., Murray J.C.;  
RT "Sequence and chromosomal assignment of human BAPX1, a bagpipe-related  
gene, to 4p16.1: a candidate gene for skeletal dysplasia";  
RL Genomics 45:425-428(1997).  
RN [3]  
RP SEQUENCE OF 206-265 FROM N.A.  
RX MEDLINE=97398454; PubMed=9256352;  
RA Tribioli C., Frasch M., Lufkin T.;  
RT "Bapx1: an evolutionary conserved homologue of the Drosophila bagpipe  
homeobox gene is expressed in splanchic mesoderm and the embryonic  
skeleton";  
RL Mech. Dev. 65:145-162(1997).

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: Expressed in visceral mesoderm and embryonic  
skeleton.  
CC -!- SIMILARITY: Belongs to the NK-3 homeobox family.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF005260; AAC39536.1; -;  
EMBL; AF009801; AAB82783.1; -;  
EMBL; AF009802; AAB82784.1; -;  
EMBL; U89845; AAB49696.1; -;  
HSSP; P22808; 1VND.  
TRANSFAC; T02668; -;  
Genew; HGNC:951; BAPX1.  
MIM; 602183; -;  
DR GO; GO:0003702; P:RNA polymerase II transcription factor acti. . . ; TAS.  
DR GO; GO:0001501; P:skeletal development; TAS.  
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR000047; HTH\_lambrepresr.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR PRINTS; PR00031; HTHREPRESSR.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Nuclear protein.  
FT DOMAIN 180 188 POLY-GLY.



FT DOMAIN 194 198 POLY-GLU.  
 FT DNA BIND 206 265 HOMEBOX.  
 FT DOMAIN 327 330 POLY-ALA.  
 SQ SEQUENCE 333 AA; 34813 MW; 8C406E188D27780B CRC64;

Query Match 62.7%; Score 52; DB 1; Length 333;  
 Best Local Similarity 62.5%; Pred. No. 1.3;  
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAPGSTA 16  
 | : ||| : ||||| |||  
 Db 25 GGLAPEGKRPAPGGTA 40

RESULT 4  
 ANCA\_CLOTH STANDARD; PRT; 447 AA.  
 AC Q06848;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cellulosome anchoring protein precursor.  
 GN ANCA.  
 OS Clostridium thermocellum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 CX NCBI\_TaxID=1515;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIB 10682;  
 RX MEDLINE=93209931; PubMed=8458832;  
 RA Fujino T., Beguin P., Aubert J.-P.;  
 RT "Organization of a Clostridium thermocellum gene cluster encoding the  
 RT cellulosomal scaffolding protein Cipa and a protein possibly involved  
 RT in attachment of the cellulosome to the cell surface."  
 RL J. Bacteriol. 175:1891-1899(1993).  
 CC -!- FUNCTION: ANCHORS THE CELLULOSE TO THE CELL SURFACE BY BINDING  
 CC THE DUPLICATED SEGMENT THAT IS PRESENT AT THE C-TERMINAL END OF  
 CC CIPA.  
 CC -!- SUBCELLULAR LOCATION: Cell wall.  
 CC -!- SIMILARITY: Contains 1 cohesin domain.  
 CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X67506; CAA47843.2; -.  
 DR PIR; T18264; T18264.  
 DR HSSP; Q06851; LAOH.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR002102; Cohesin.  
 DR InterPro; IPR001119; SLH.  
 DR Pfam; PF00963; Cohesin; 1.  
 DR Pfam; PF00395; SLH; 3.  
 DR PROSITE; PS01072; SLH\_DOMAIN; 2.  
 DR Cellulose degradation; Cell wall; S-layer; Signal; Repeat.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT CHAIN 30 447 CELLULOSE ANCHORING PROTEIN.  
 FT DOMAIN 30 180 RECEPTOR BINDING SITE FOR DUPLICATED  
 FT SEGMENT OF CIPA (POTENTIAL).  
 FT DOMAIN 30 180 COHESIN.  
 FT DOMAIN 181 237 GLY/PRO/SER/THR-RICH.  
 FT DOMAIN 241 283 SLH 1 (INCOMPLETE).  
 FT DOMAIN 284 348 SLH 2.  
 FT DOMAIN 349 408 SLH 3.  
 FT DOMAIN 409 429 SLH 4 (INCOMPLETE).  
 SQ SEQUENCE 447 AA; 48530 MW; 0FD674134ABB8FE1 CRC64;

Query Match 59.0%; Score 49; DB 1; Length 447;  
 Best Local Similarity 69.2%; Pred. No. 5;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 SAPDTREAPGSTA 16  
 ||| : |||||  
 Db 176 SAPEARTPGSTA 188

RESULT 5  
 ZYX\_MOUSE STANDARD; PRT; 564 AA.  
 AC Q62523; P70461;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Zyxin.  
 GN ZYX.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR X Swiss Webster;  
 RX MEDLINE=97094926; PubMed=8940160;  
 RA Macalima T., Otte J., Hensler M.E., Bockholt S.M., Louis H.A.,  
 RA Kalif-Suske M., Grzeschik K.H., von der Ahe D., Beckerle M.C.;  
 RT "Molecular characterization of human zyxin."  
 RL J. Biol. Chem. 271:31470-31478(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Otte J., Heischmann A., Breier G., Beckerle M.C., von der Ahe D.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Adhesion plaque protein. Binds alpha-actinin and the CRP  
 CC protein. May be a component of a signal transduction pathway that  
 CC mediates adhesion-stimulated changes in gene expression (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; associates with the actin  
 CC cytoskeleton near the adhesion plaques.  
 CC -!- SIMILARITY: Contains 3 LIM zinc-binding domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Y07711; CAA68984.1; -.  
 DR EMBL; X99063; CAA67510.1; -.  
 DR MGD; MGI:103072; Zyx.  
 DR InterPro; IPR001781; LIM.  
 DR Pfam; PF00412; LIM; 3.  
 DR ProDom; PD000094; LIM; 3.  
 DR SMART; SM00132; LIM; 3.  
 DR PROSITE; PS00478; LIM\_DOMAIN\_1; 2.  
 DR PROSITE; PS0023; LIM\_DOMAIN\_2; 3.  
 KW Repeat; LIM domain; Metal-binding; Zinc; Cell adhesion.  
 FT DOMAIN 64 77 PRO-RICH.  
 FT DOMAIN 94 138 PRO-RICH.  
 FT DOMAIN 376 435 LIM 1.  
 FT DOMAIN 436 495 LIM 2.  
 FT DOMAIN 496 562 LIM 3.  
 FT CONFLICT 215 215 R -> A (IN REF. 1).  
 FT CONFLICT 284 292 IKKWCRLMP -> NQKWPVPPDA (IN REF. 1).  
 FT CONFLICT 484 484 S -> C (IN REF. 1).  
 SQ SEQUENCE 564 AA; 60790 MW; 001E1B3C82ADA1EB CRC64;

Query Match 56.6%; Score 47; DB 1; Length 564;  
 Best Local Similarity 72.7%; Pred. No. 13;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 PDTRPAPGSTA 16  
|:|||||  
Db 186 PSTKPAPGGTA 196

## RESULT 6

CUBB HUMAN  
ID CU88 HUMAN STANDARD; PRT; 145 AA.  
AC P59052;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Protein C21orf88.  
GN C21orf88.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=22032984; PubMed=12036297;  
RA Raymond A., Camargo A.A., Deutsch S., Stevenson B.J., Parmigiani R.B.,  
RA Ucla C., Bettoni F., Rossier C., Lyle R., Guipponi M., de Souza S.,  
RA Iseli C., Jongeneel C.V., Bucher P., Simpson A.J.G., Antonarakis S.E.;  
RT "Nineteen additional unreported transcripts from human chromosome  
21.";  
RL Genomics 79:824-832(2002).

## -!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=1;

Isoid=P59052-1; Sequence=Displayed;

Name=2;

Isoid=P59052-2; Sequence=VSP\_003836, VSP\_003837;

Note=No experimental confirmation available;

-!- TISSUE SPECIFICITY: Ubiquitous.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; AF426266; AAM53522.1; -;  
Genew; HGNC:16424; C21orf88.

Alternative splicing.

VARSPLOC 52 64 GGQKSHGLCWLLC -> NIQWHRDQLGDN (in  
isoform 2).

VARSPLOC 65 145 /FTID=VSP\_003836.  
Missing (in isoform 2).

FT SEQUENCE 145 AA; 15789 MW; DDE0023065CE0C51 CRC64;

Query Match 55.4%; Score 46; DB 1; Length 145;  
Best Local Similarity 57.1%; Pred. No. 4.3;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAPGS 14

Db 28 GCTTAPQQRPPPGT 41  
|:|||||

## RESULT 7

DIAL\_MOUSE  
ID DIAL\_MOUSE STANDARD; PRT; 1255 AA.  
AC O08808;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Diaphanous protein homolog 1 (Diaphanous-related formin 1) (DRF1)

(mDIA1) (p140mDIA).  
DIAPH1 OR DIAP1.  
Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97357293; PubMed=9214622;  
RA Watanabe N., Madaule P., Reid T., Ishizaki T., Watanabe G.,  
RA Kakizuka A., Saito Y., Nakao K., Jockusch B.M., Narumiya S.;  
RT "p140mDIA, a mammalian homolog of Drosophila diaphanous, is a target  
protein for Rho small GTPase and is a ligand for profilin.";  
RL EMO J. 16:3044-3056(1997).  
RN [2]  
RP FUNCTION.

RX MEDLINE=20142655; PubMed=10678165;  
RA Tomimaga T., Sahai E., Chardin P., McCormick F., Courtneidge S.A.,  
RA Alberts A.S.;  
RT "Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase  
signaling.";  
RL Mol. Cell 5:13-25(2000).

-!- FUNCTION: Binds to GTP-bound form of Rho and to profilin. Acts in  
a Rho-dependent manner to recruit profilin to the membrane, where  
it promotes actin polymerization. It is required for cytokinesis,  
stress fiber formation, and transcriptional activation of the  
serum response factor. DFR proteins couple Rho and Src tyrosine  
kinase during signaling and the regulation of actin dynamics.

-!- SUBCELLULAR LOCATION: MEMBRANE RUFFLES, ESPECIALLY AT THE TIP OF  
RUFFLES, OF MOTILE CELLS.

-!- TISSUE SPECIFICITY: Ubiquitous.

-!- DOMAIN: DRPs are regulated by intramolecular GBD-DAD binding where  
Rho-GTP activates the DRFs by disrupting the GBD-DAD interaction.

-!- SIMILARITY: Contains 1 GTPase-binding (GBD) domain.

-!- SIMILARITY: Contains 1 Formin homology 1 (FH1) domain.

-!- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.

-!- SIMILARITY: Contains 1 Formin homology 3 (FH3) domain.

-!- SIMILARITY: Contains 1 DRF autoregulatory (DAD) domain.

-!- SIMILARITY: Belongs to the formin homology family. Diaphanous  
subfamily.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U96963; AAC53280.1; -;  
PIR; T31065; T31065.

MGI; MGI:1194490; Diapl.

GO; GO:0005515; F:protein binding; IPI.

InterPro; IPR003104; FH2.

Pfam; PF02181; FH2; 1.

SMART; SM00498; FH2; 1.

Coiled coil; Repeat.

DOMAIN 460 562 COILED COIL (POTENTIAL).

DOMAIN 63 260 GBD.

DOMAIN 157 457 FH3.

DOMAIN 586 747 FH1 (PRO-RICH).

DOMAIN 752 1197 FH2.

DOMAIN 1027 1179 COILED COIL (POTENTIAL).

DOMAIN 1180 1194 DAD.

DOMAIN 1196 1199 ARG/LYS-RICH (BASIC).

SEQUENCE 1255 AA; 139343 MW; 09404164873CA7C1 CRC64;

Query Match 54.2%; Score 45; DB 1; Length 1255;  
Best Local Similarity 56.2%; Pred. No. 57;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAPGSTA 16

||| ||| ||| ||| |||

```
Db 643 GVASIPPP2PLPGATA 658
RESULT 8
NK32 MOUSE
ID NK32_MOUSE STANDARD; PRT; 333 AA.
AC P97503;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein Nkx-3.2 (Bagpipe homeobox protein homolog 1).
GN BAPX1 OR NKX3B OR NKX3-2 OR NKX-3.2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=97398454; PubMed=9256352;
RA Tribioli C., Frasch M., Lufkin T.;
RT Bapx1: an evolutionary conserved homologue of the Drosophila bagpipe
RT homeobox gene is expressed in splanchic mesoderm and the embryonic
RT skeleton."
RL Mech. Dev. 55:145-162(1997).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Expressed in visceral mesoderm and embryonic
CC skeleton.
CC -!- SIMILARITY: Belongs to the NK-3 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U87957; AAB47764.1; -.
CC HSP; P22808; 1VND.
CC TRANSFAC; T02667; -.
CC MGD; MGI:108015; Bapx1.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR000047; HTH_lambrepresr.
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC PRINTS; PR00031; HTHREPRESSR.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS50071; HOMEBOX_2; 1.
CC DNA-binding; Homeobox; Nuclear protein.
FT DOMAIN 194 198 POLY-GLU.
FT DNA_BIND 206 265 HOMEBOX.
FT DOMAIN 327 330 POLY-ALA.
SQ SEQUENCE 333 AA; 35192 MW; 1475D60C0774F226 CRC64;

Query Match 53.0%; Score 44; DB 1; Length 333;
Best Local Similarity 53.3%; Pred. No. 20;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAGST 15
| : |||||
Db 25 GGLATPEGRPAGGT 39

RESULT 9
FX11 HUMAN
ID FX11_HUMAN STANDARD; PRT; 351 AA.
AC Q12951; Q14518; Q8N6L8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAY-2004 (Rel. 43, Last annotation update)
```

Forkhead box protein 11 (Forkhead-related protein FKHL10) (Forkhead-related transcription factor 6) (FREAC-6) (Hepatocyte nuclear factor 3 forkhead homolog 3) (HNF-3/fork-head homolog-3) (HFH-3).

FOX11 OR FKHL10 OR FREAC6.

Homo sapiens (Human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI\_TaxID=9606;

[1]

SEQUENCE FROM N.A. (ISOFORM 1).

TISSUE=Kidney;

MEDLINE=97298077; PubMed=9153225;

Ozerdier D.G., Ye H., Peterson R.S., Clevidence D.E., Costa R.H.;

"The winged helix transcriptional activator HFH-3 is expressed in the distal tubules of embryonic and adult mouse kidney."

J. Biol. Chem. 272:13725-13730(1997).

[2]

SEQUENCE FROM N.A. (ISOFORM 2).

TISSUE=Colon, and Kidney;

MEDLINE=22388257; PubMed=12477932;

Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usslin T.B., Toshitoki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[3]

SEQUENCE OF 91-166 FROM N.A.

MEDLINE=95045392; PubMed=7957066;

Pierrou S., Hellqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;

"Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending."

EMBO J. 13:5002-5012(1994).

-!- FUNCTION: Transcription activator. May act on the genes for Na/K-ATPase, Na/H and anion exchangers, E-cadherin and mineralocorticoid receptor as well as on the genes for transcription factors HNF-1, VHN-1 and HNF-4.

-!- SUBCELLULAR LOCATION: Nuclear.

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=1;

IsoId=Q12951-1; Sequence=Displayed;

Name=2;

IsoId=Q12951-2; Sequence=VSP\_001543;

Note=No experimental confirmation available;

TISSUE SPECIFICITY: Kidney specific.

-!- SIMILARITY: Contains 1 fork-head domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; L13203; AAB50574.1; -.
DR EMBL; BC029773; AAB50574.1; -.
DR EMBL; U13224; AAA92041.1; -.
DR

```

DR PIR; S51629; S51629.
DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T02474; -.
DR Genew; HGNC:3815; FOX11.
DR MIM; 601093; -.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0008301; F:DNA bending activity; NAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007275; P:development; NAS.
DR GO; GO:0005355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
DR DNA-binding; Nuclear protein; Transcription regulation; Activator;
KW Alternative splicing.
FT DNA BIND 95 186 FORK-HEAD.
FT VARSPIC 165 259 Missing (in isoform 2).
FT SEQUENCE 351 AA; 38091 MW; 31198DC203913D51 CRC64;
SQ

Query Match 53.0%; Score 44; DB 1; Length 351;
Best Local Similarity 58.3%; Pred.No. 21;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVTSPADTRPAP 12
Db | | | | | | | | | |
232 GTSSPEKRRSP 243

RESULT 10
AMPM MANSE
ID AMPM MANSE STANDARD; PRT; 990 AA.
AC Q11001;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Membrane alanyl aminopeptidase precursor (EC 3.4.11.-)
DE (Aminopeptidase N-like protein) (CRYIA(C) receptor) (Fragment).
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sphingioidea;
OC Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Midgut;
RX MEDLINE=9535365; PubMed=7629076;
RA Knight P.J.K., Knowles B.H., Ellar D.J.;
RT "Molecular cloning of an insect aminopeptidase N that serves as a
receptor for Bacillus thuringiensis CryIA(c) toxin.";
RL J. Biol. Chem. 270:17765-17770(1995).
CC -!- FUNCTION: Binds to the B.thuringiensis toxin, CRYIA(C).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- TISSUE SPECIFICITY: Midgut brush-border membrane.
CC -!- SIMILARITY: Belongs to peptidase family M1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X89081; CAA61452.1; -.
CC EMBL; AB007039; BAA32476.1; -.
CC MEROPS; M01.013; -.

```

```

DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001930; Peptidase_M1.
DR Pfam; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALADIPITASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc; Glycoprotein;
KW GPI-anchor; Signal; Lipoprotein.
FT NON_TER 1 1
FT SIGNAL <1 15 POTENTIAL.
FT PROPEP 16 35 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 36 968 MEMBRANE ALANYL AMINOPEPTIDASE.
FT PROPEP 969 990 REMOVED IN MATURE FORM (POTENTIAL).
FT METAL 357 357 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 358 358 BY SIMILARITY.
FT METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 380 380 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 460 460 PROTON DONOR (POTENTIAL).
FT LIPID 968 968 GPI-anchor amidated glycine (Potential).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 609 609 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 990 AA; 111293 MW; EDF6E63C398C11D8 CRC64;

Query Match 53.0%; Score 44; DB 1; Length 990;
Best Local Similarity 75.0%; Pred.No. 62;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGS 14
Db | | | | | | | | | |
958 TVAPATTPAPGS 969

RESULT 11
PCPR SPCHR
ID PCPR SPCHR STANDARD; PRT; 303 AA.
AC P52679;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE PCP degradation transcriptional activation protein.
GN PCPR.
OS Sphingobium chlorophenolicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Sphingomonadaceae; Sphingobium.
OX NCBI_TaxID=46429;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 39723;
RA Lange C.C., Orser C.S.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transcriptional activator for the pcg genes for
pentachlorophenol degradation.
CC -!- SIMILARITY: Contains 1 HTH lysr-type DNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U12290; AAA68939.1; -.
DR InterPro; IPR000847; HTH_LysR.
DR Pfam; PF00126; HTH_1; 1.
DR PROSITE; PS50931; HTH_LYSR; 1.
KW Aromatic hydrocarbons catabolism; Transcription regulation;
KW DNA-binding; Activator.
FT DOMAIN 5 62 HTH_LYSR-TYPE.
FT DNA BIND 22 41 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 303 AA; 33549 MW; 471A1DD3FA345C00 CRC64;

```



```
Query Match      51.8%; Score 43; DB 1; Length 303;
Best Local Similarity 58.3%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 VTSAPDTRPAPG 13
Db 174 IPRPDARPGPG 185

RESULT 12
CIWG HUMAN
ID CIWG HUMAN STANDARD; PRT; 309 AA.
AC Q96T55; Q9H591;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Potassium channel subfamily K member 16 (TWIK-related alkaline pH
DE activated K+ channel 1) (2P domain potassium channel Talk-1).
GN KCNK16 OR TALK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=21164727; PubMed=11263999;
RA Girard C., Duprat F., Terrenoire C., Tinel N., Fosset M., Romey G.,
RA Lazdunski M., Lesage F.;
RT "Genomic and functional characteristics of novel human pancreatic 2P
RT domain K(+) channels."
RL Biochem. Biophys. Res. Commun. 282:249-256(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Williams S.;
RL Submitted (SBP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Outward rectifying potassium channel. Produces rapidly
CC activating and non-inactivating outward rectifier K(+) currents.
CC -!- SUBUNIT: Homodimer (Potential).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q96T55-1; Sequence=Displayed;
CC IsoId=Q96T55-2; Sequence=VSP_006699;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in pancreas. Not detectable
CC in the other tissues tested.
CC -!- MISCELLANEOUS: Inhibited by Ba(+2), quinine, quinidine, chloroform
CC and halothane. Activated at alkaline pH.
CC -!- SIMILARITY: Belongs to the two pore domain potassium channel
CC family.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF358909; AAK49532.1; -
EMBL; AL136087; CAC07336.1; -
Genew; HGNC:14464; KCNK16.
MIM; 607369; -
InterPro; IPR003280; K+channel_2pore.
InterPro; IPR001622; K+channel_pore.
InterPro; IPR003092; TASK channel.
PRINTS; PR01333; 2PORECHANNEL.
PRINTS; PR01095; TASKCHANNEL.
KW Transport; Ion transport; Ionic channel; Voltage-gated channel;
KW Potassium channel; Potassium; Transmembrane; Alternative splicing.
```

```
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL.
FT DOMAIN 98 116 PORE-FORMING 1 (POTENTIAL).
FT TRANSMEM 120 140 POTENTIAL.
FT DOMAIN 141 165 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 166 186 POTENTIAL.
FT DOMAIN 202 221 PORE-FORMING 2 (POTENTIAL).
FT TRANSMEM 238 258 POTENTIAL.
FT DOMAIN 259 309 CYTOPLASMIC (POTENTIAL).
FT VARSPLIC 269 309 LRQCGAKAAGRRRPRRGSTAARGVQVTPQDFPISKGLGS
FT -> RGLGVKDGASDPSGLPRPQKIPIA (in isoform
FT 2).
SQ SEQUENCE 309 AA; 34153 MW; 99C4B11EB26B0764 CRC64;
/FTId=VSP_006699.
Query Match 51.8%; Score 43; DB 1; Length 309;
Best Local Similarity 56.2%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAPGSTA 16
Db 274 GAKAAPGRRRPRRGSTA 289

RESULT 13
MID2 YEAST
ID MID2 YEAST STANDARD; PRT; 376 AA.
AC P36027; P41944;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mating process protein MID2 (Serine-rich protein SMS1) (Protein kinase
DE A interference protein).
GN MID2 OR SMS1 OR KAI1 OR YLR332W OR L8543.19.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Sukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95129360; PubMed=7828875;
RA Ono T., Suzuki T., Anraku Y., Iida H.;
RT "The MID2 gene encodes a putative integral membrane protein with a
RT Ca(2+)-binding domain and shows mating pheromone-stimulated
RT expression in Saccharomyces cerevisiae."
RL Gene 151:203-208(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 70036;
RX MEDLINE=95092800; PubMed=7999801;
RA Takeuchi J., Okada M., Toh-E A., Kikuchi Y.;
RT "The SMS1 gene encoding a serine-rich transmembrane protein
RT suppresses the temperature sensitivity of the htr1 disruptant in
RT Saccharomyces cerevisiae."
RL Biochim. Biophys. Acta 1260:94-96(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Killier L., Riles L., Albermann K., Andre B., Ansoerge M.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII."
RL Nature 387:87-90(1997).
RN [4]
```

RP SEQUENCE OF 60-304 FROM N.A.  
 RX MEDLINE=93360904; PubMed=8355657;  
 RA Daniel J.;  
 RT "Potentially rapid walking in cellular regulatory networks using the  
 gene-gene interference method in yeast.";  
 RL Mol. Gen. Genet. 240:245-257(1993).  
 CC -!- FUNCTION: Seems to be involved in start control. May have a  
 possible role in the trapping of phosphorylation somehow linked to  
 the start I control.  
 CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).  
 CC -!- SIMILARITY: TO YEAST YGR23W.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D29945; BAA06220.1; -;  
 DR EMBL; D29964; BAA06230.1; -;  
 DR EMBL; U20618; AAB64527.1; -;  
 DR PIR; S52137; S52137.  
 DR GermOnline; 142396; -;  
 DR SGD; S0004324; MID2.  
 DR GO; GO:0005887; C: integral to plasma membrane; IDA.  
 DR GO; GO:0004888; F: transmembrane receptor activity; IGI.  
 DR GO; GO:0007047; P: cell wall organization and biogenesis; IGI.  
 DR GO; GO:0000767; P: cellular morphogenesis during conjugation; IGI.  
 DR InterPro; IPR007567; MID2.  
 DR Pfam; PF04478; MID2; 1.  
 FT DOMAIN 31 184 SER/THR-RICH.  
 FT CONFLICT 303 304 GG -> VV (IN REF. 4).  
 SQ SEQUENCE 376 AA; 39146 MW; 3612E260BC1B04A2 CRC64;

Query Match 51.8%; Score 43; DB 1; Length 376;  
 Best Local Similarity 53.3%; Pred. No. 32;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VTSAPDTRPAPGSTA 16  
 :||| | :||  
 Db -61 ITSAPSTSTPSTTA 175  
 RESULT 14  
 ODP2 MYCPN  
 ID ODP2 MYCPN STANDARD; PRT; 402 AA.  
 AC P75392;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase  
 DE complex (EC 2.3.1.12) (E2).  
 GN PDHC OR MPN391 OR MP447.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
 RA Herrmann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 pneumoniae.";  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL  
 CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE  
 COPIES OF THREE ENZYMIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),  
 DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE  
 (E3) (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide = CoA + S-

CC acetyldihydrolipoamide.  
 CC -!- COFACTOR: Contains 1 covalently bound lipoyl cofactor (By  
 CC similarity).  
 CC -!- SUBUNIT: Forms a 24-polypeptide structural core with octahedral  
 CC symmetry (By similarity).  
 CC -!- SIMILARITY: Belongs to the 2-oxoacid dehydrogenase family.  
 CC -!- SIMILARITY: Contains 1 lipoyl-binding domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000044; AAB96095.1; -;  
 DR PIR; S73773; S73773.  
 DR HSSP; P11961; ILAB.  
 DR InterPro; IPR001078; 2Oxoacid dh.  
 DR InterPro; IPR000089; Biotin\_lipoyl.  
 DR InterPro; IPR003016; Lipoyl\_BS.  
 DR Pfam; PF00198; 2-oxoacid dh; 1.  
 DR Pfam; PF00364; Biotin\_lipoyl; 1.  
 DR ProDom; PD001115; 2Oxoacid dh; 1.  
 DR PROSITE; PS00189; LIPOYL; 1.  
 KW Glycolysis; Transferase; Acyltransferase; Lipoyl; Complete proteome.  
 FT BINDING 43 43 LIPOYL (BY SIMILARITY).  
 FT ACT\_SITE 374 374 POTENTIAL.  
 SQ SEQUENCE 402 AA; 42397 MW; F09314A9E714A1D6 CRC64;

Query Match 51.8%; Score 43; DB 1; Length 402;  
 Best Local Similarity 64.3%; Pred. No. 34;  
 Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGSTA 16  
 :||| | :|||  
 Db 154 TSAPDTRPAPASAA 167

RESULT 15  
 BLT4 HORVU  
 ID BLT4 HORVU STANDARD; PRT; 130 AA.  
 AC P25307;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE BLT4 protein precursor.  
 GN BLT4.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Igri; TISSUE=Meristem;  
 RX MEDLINE=92049237; PubMed=1944226;  
 RA Dunn M.A., Hughes M.A., Zhang L., Pearce R.S., Quigley A.S.,  
 RA Jack P.L.;  
 RT "Nucleotide sequence and molecular analysis of the low temperature  
 induced cereal gene, BLT4.";  
 RL Mol. Gen. Genet. 229:389-394(1991).  
 CC -!- FUNCTION: Possible dehydrative stress responsive protein. Not  
 shown to have lipid transfer activity.  
 CC -!- TISSUE SPECIFICITY: Shoot meristem.  
 CC -!- INDUCTION: By low temperature and drought.  
 CC -!- SIMILARITY: Belongs to the plant LTP family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X56547; CAA39887.1; -.  
PIR; S17961; S17961.  
HSSP; P24296; LBWO.  
InterPro; IPR003612; AAI.  
InterPro; IPR000528; Plant\_LTP.  
Pfam; PF00234; tryo\_alpha\_amyl; 1.  
PRINTS; PR00382; LIPIDTRNSFER.  
SMART; SM00499; AAI; 1.  
PROSITE; PS00597; PLANT\_LTP; FALSE\_NDG.  
Signal.  
SIGNAL 1 25 POTENTIAL.  
CHAIN 26 130 BLT4 PROTEIN.  
SEQUENCE 130 AA; 13560 MW; 8DB7ED2F99219EE0 CRC64;

Query Match 50.6%; Score 42; DB 1; Length 130;  
Best Local Similarity 62.5%; Pred. No. 15;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GVTSAPDTRPAPGSTA 16  
Db 95 GAASAPSTRSAPVSTA 110

RESULT 16  
EVL\_RAT  
ID EVL\_RAT STANDARD; PRT; 393 AA.  
AC O08719;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ena/vasodilator stimulated phosphoprotein-like protein (Ena/VASP-like protein).  
DE EVL OR RNB6.  
OS Rattus norvegicus (Rat).  
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]\_TaxID=10116;  
RP SEQUENCE FROM N.A.; TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RC STRAIN=Wistar; TISSUE=Brain;  
RX MEDLINE=97415794; PubMed=9268706;  
RA Ohta S., Mineta T., Kimoto M., Tabuchi K.;  
RT "Differential display cloning of a novel rat cDNA (RNB6) that shows high expression in the neonatal brain revealed a member of Ena/VASP family.";  
RT Biochem. Biophys. Res. Commun. 237:307-312(1997).  
CC -!- FUNCTION: Enhances actin nucleation and polymerization (By similarity).  
CC -!- SUBUNIT: Binds to the SH3 domains of ABL1, LYN and SRC. Also binds to profilin and the WW domain of APBB1/FE55. Binds to SEMA6A (By similarity).  
CC -!- SUBCELLULAR LOCATION: Focal adhesions (By similarity).  
CC -!- TISSUE SPECIFICITY: Expression detected in brain, spleen, thymus and testis.  
CC -!- DEVELOPMENTAL STAGE: In the brain, expression gradually increases during embryonic development, reaches a maximum at postnatal day 1 and decreases thereafter.  
CC -!- PTM: Phosphorylated by PKA; phosphorylation abolishes binding to SH3 domains of ABL and SRC (By similarity).  
CC -!- SIMILARITY: Contains 1 WH1 domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U70211; AAC53322.1; -.  
DR PIR; JC5614; JC5614.  
DR GO; GO:0005737; C:cytoplasm; ISS.  
DR GO; GO:0005925; C:focal adhesion; ISS.  
DR GO; GO:0030027; C:lamellipodium; ISS.  
DR GO; GO:0008580; F:cytoskeletal regulator activity; ISS.  
DR GO; GO:0003782; F:actin capping activity; ISS.  
DR GO; GO:0005522; F:profilin binding; ISS.  
DR GO; GO:0017124; F:SH3-domain binding; ISS.  
DR GO; GO:0030048; P:actin filament-based movement; ISS.  
DR GO; GO:0045010; P:actin nucleation; ISS.  
DR GO; GO:0008154; P:actin polymerization and/or depolymerization; ISS.  
DR GO; GO:0007411; P:axon guidance; NAS.  
DR GO; GO:0006928; P:cell motility; NAS.  
DR GO; GO:0007166; P:cell surface receptor linked signal transdu. .; ISS.  
DR GO; GO:0007417; P:central nervous system development; NAS.  
DR GO; GO:0009887; P:organogenesis; ISS.  
DR GO; GO:0030168; P:platelet activation; ISS.  
DR InterPro; IPR000697; EVH1.  
DR InterPro; IPR001960; WH1.  
DR Pfam; PF00568; WH1; 1.  
DR SMART; SM00461; WH1; 1.  
KW SH3-binding; Phosphorylation.  
FT DOMAIN 1 109 WH1.  
FT DOMAIN 160 204 PRO-RICH.  
SQ SEQUENCE 393 AA; 42095 MW; 6371D91362925D4E CRC64;

Query Match 50.6%; Score 42; DB 1; Length 393;  
Best Local Similarity 57.1%; Pred. No. 47;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TSAPDTRPAPGSTA 16  
Db 296 TEDPSTSPSPSGRA 309

RESULT 17  
PAXI CHICK  
ID PAXI CHICK STANDARD; PRT; 559 AA.  
AC P49024;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Paxillin.  
DE PAXN.  
OS Gallus gallus (Chicken).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
CC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]\_TaxID=9031;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95197488; PubMed=7534286;  
RA Salgia R., Li J.-L., Lo S.H., Brunkhorst B., Kansas G.S.,  
RA Sobhany E.S., Sun Y., Pisick E., Hallek M., Ernst T., Tantravahi R.,  
RA Chen L.B., Griffin J.D.;  
RT "Molecular cloning of human paxillin, a focal adhesion protein phosphorylated by P210BCR/ABL.";  
RL J. Biol. Chem. 270:5039-5047(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95051116; PubMed=7525621;  
RA Turner C.E., Miller J.T.;  
RT "Primary sequence of paxillin contains putative SH2 and SH3 domain binding motifs and multiple LIM domains: identification of a vinculin and p125Fak-binding region.";  
RL J. Cell Sci. 107:1583-1591(1994).  
RN [3]  
RP PHOSPHORYLATION.  
RX MEDLINE=95340539; PubMed=7615549;  
RA Bellis S.L., Miller J.T., Turner C.E.;  
RT "Characterization of tyrosine phosphorylation of paxillin in vitro by focal adhesion kinase.";

RL J. Biol. Chem. 270:17437-17441(1995).

CC -!- FUNCTION: Cytoskeletal protein involved in actin-membrane

CC attachment at sites of cell adhesion to the extracellular matrix

CC (focal adhesion). Binds in vitro to vinculin as well as to the SH3

CC domain of c-SRC and, when tyrosine phosphorylated, to the SH2

CC domain of v-CRK.

CC -!- PTM: Phosphorylated on tyrosine residues during integrin-mediated

CC cell adhesion, embryonic development, fibroblast transformation

CC and following stimulation of cells by mitogens.

CC -!- SIMILARITY: Contains 4 LIM zinc-binding domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U14589; AAC59665.1; -.

DR EMBL; L30099; AAC38018.1; -.

DR PIR; B55933; B55933.

DR HSSP; P04006; LIML.

DR InterPro; IPR001781; LIM.

DR InterPro; IPR001904; Paxillin.

DR Pfam; PF00412; LIM; 4.

DR Pfam; PF03535; Paxillin; 1.

DR PRINTS; PR00832; PAXILLIN.

DR ProDom; PD000094; LIM; 4.

DR SMART; SM00132; LIM; 4.

DR PROSITE; PS00478; LIM\_DOMAIN\_1; 4.

DR PROSITE; PS00023; LIM\_DOMAIN\_2; 4.

DR Cytoskeleton; Phosphorylation; LIM domain; Repeat; Metal-binding;

DR Zinc.

FT DOMAIN 46 53 PRO-RICH.

FT DOMAIN 326 376 LIM 1.

FT DOMAIN 385 435 LIM 2.

FT DOMAIN 444 494 LIM 3.

FT DOMAIN 503 553 LIM 4.

FT MOD\_RES 31 31 PHOSPHORYLATION (BY SIMILARITY).

FT MOD\_RES 118 118 PHOSPHORYLATION (BY FAK1).

FT SEQUENCE 559 AA; 61242 MW; 6450270D90B2DE84 CRC64;

QY 1 GVTSAPDTRPAGS 14

Db 287 GSSSPSTTPKPGS 300

Query Match 50.6%; Score 42; DB 1; Length 559;

Best Local Similarity 57.1%; Pred. No. 68;

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 GVTSAPDTRPAGS 14

Db 287 GSSSPSTTPKPGS 300

RESULT 18

SR72\_CANFA

ID SR72 CANFA STANDARD; PRT; 670 AA.

AC P33731;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Signal recognition particle 72 kDa protein (SRP72).

GN SRP72.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI\_TaxID=9615;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=93273803; PubMed=8388879;

RA Luetcke H., Prehn S., Ashford A.J., Remus M., Frank R.,

RA Dobberstein B.;

RT "Assembly of the 68- and 72-kD proteins of signal recognition

RT particle with 7S RNA."

RL J. Cell Biol. 121:977-985(1993).

CC -!- FUNCTION: Signal-recognition-particle assembly has a crucial role

CC in targeting secretory proteins to the rough endoplasmic

CC reticulum membrane. SRP72 binds the 7S RNA only in presence of

CC SRP68. This ribonucleoprotein complex might interact directly with

CC the docking protein in the ER membrane and possibly participate

CC in the elongation arrest function.

CC -!- SUBUNIT: Signal recognition particle consists of a 7S RNA molecule

CC of 300 nucleotides and six protein subunits: SRP72, SRP68, SRP54,

CC SRP19, SRP14 and SRP9.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- DOMAIN: The C-terminus is essential for the interaction with the

CC SRP68/7S RNA complex.

CC -!- SIMILARITY: Belongs to the SRP72 family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X67813; CAA48014.1; -.

DR PIR; A40692; A40692.

DR InterPro; IPR008941; TPR-like.

DR InterPro; IPR001440; TPR.

DR Pfam; PF00515; TPR; 2.

DR Signal recognition particle; Ribonucleoprotein.

FT INIT\_MET 0 0 PROBABLE.

FT MOD\_RES 1 1 BLOCKED.

FT DOMAIN 551 560 POLY-LYS.

FT DOMAIN 661 664 POLY-LYS.

FT SEQUENCE 670 AA; 74362 MW; D394CC56600B5C3D CRC64;

QY 4 SAPDTRPAGSTA 16

Db 619 SSPTSPRPGSAA 631

Query Match 50.6%; Score 42; DB 1; Length 670;

Best Local Similarity 61.5%; Pred. No. 82;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 SAPDTRPAGSTA 16

Db 619 SSPTSPRPGSAA 631

RESULT 19

SR72\_HUMAN

ID SR72 HUMAN STANDARD; PRT; 670 AA.

AC Q76094;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Signal recognition particle 72 kDa protein (SRP72).

GN SRP72.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Gowda K., Zwiab C.;

RT "Protein SRP72 sequence of human signal recognition particle.";

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Utz P.J., Hottel M., Miller I.J., Anderson P.;

RT "Sequence of human signal recognition particle (SRP) 72.";

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Signal-recognition-particle assembly has a crucial role

CC in targeting secretory proteins to the rough endoplasmic

CC reticulum membrane. SRP72 binds the 7S RNA only in presence of

CC SRP68. This ribonucleoprotein complex might interact directly with

CC the docking protein in the ER membrane and possibly participate

CC in the elongation arrest function.

CC -!- SUBUNIT: Signal recognition particle consists of a 7S RNA molecule

```

CC of 300 nucleotides and six protein subunits: SRP72, SRP68, SRP54,
CC SRP19, SRP14 and SRP9.
CC
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DOMAIN: The C-terminus is essential for the interaction with the
CC SRP68/7S RNA complex (By similarity).
CC -!- SIMILARITY: Belongs to the SRP72 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF077019; AAC27324.1; -.
CC EMBL; AF069765; AAC97490.1; -.
CC Genew; HGNC:11303; SRP72.
CC MIM; 602122; -.
CC InterPro; IPR008941; TPR-like.
CC InterPro; IPR001440; TPR.
CC Pfam; PF00515; TPR; 2.
CC Signal recognition particle; Ribonucleoprotein.
CC INIT MET 0 BY SIMILARITY.
CC DOMAIN 551 560 POLY-LYS.
CC DOMAIN 661 664 POLY-LYS.
CC SEQUENCE 670 AA; 74475 MW; FBOF7F310F53FFB1 CRC64;
CC
CC Query Match 50.6%; Score 42; DB 1; Length 670;
CC Best Local Similarity 61.5%; Pred. No. 82;
CC Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
CC
CC QY 4 SAPDTRPAPGSTA 16
CC Db 619 SSPTSPRPGSAA 631
CC
CC RESULT 20
CC MERK MOUSE STANDARD; PRT; 994 AA.
CC
CC AC Q69805; Q62194;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Proto-oncogene tyrosine-protein kinase MER precursor (EC 2.7.1.112)
CC DE (C-mer) (Receptor tyrosine kinase MerTK).
CC GN MERTK OR MER.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC OX NCBI_TaxID=10090;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=C57BL/6; TISSUE=Spleen;
CC RX MEDLINE=95303502; PubMed=7784083;
CC RA Graham D.K., Bowman G.W., Dawson T.L., Stanford W.L., Earp H.S.,
CC RA Snodgrass H.R.;
CC RT "Cloning and developmental expression analysis of the murine c-mer
CC RT tyrosine kinase."
CC RL Oncogene 10:2349-2359(1995).
CC RN [2]
CC RP SEQUENCE OF 472-994 FROM N.A.
CC RC STRAIN=CD-1; TISSUE=Testis;
CC RA Dowds C.A., Burks D.J., Saling P.M.;
CC RT "A cDNA encoding part of a novel putative receptor tyrosine kinase."
CC RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC CC tyrosine phosphate.
CC CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC CC -!- TISSUE SPECIFICITY: Seems to be expressed predominantly if not
CC CC exclusively in the monocytic lineage.
CC CC -!- DEVELOPMENTAL STAGE: Expressed during most, if not all, stages of
CC CC embryological development beginning in the morula and blastocyst

```

```

CC and progressing through the yolk sac and fetal liver stages.
CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
CC AXL/UFO SUBFAMILY.
CC -!- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U21301; AAA80222.1; -.
CC EMBL; L11625; AAA85355.1; -.
CC PIR; I49276; I49276.
CC HSP; P11362; LFGL.
CC MGD; MGI:96965; Mertk.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR001245; Tyr_kinase.
CC InterPro; IPR008266; Tyr_kinase_AS.
CC Pfam; PF00041; fn3; 2.
CC Pfam; PF00047; ig; 2.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Prot_kinase; 1.
CC SMART; SM00060; FN3; 2.
CC SMART; SM00409; IG; 2.
CC SMART; SM00219; TyrKC; 1.
CC PROSITE; PS00835; IG LIKE; 2.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
CC Transferase; Phosphorylation; Transmembrane; Signal; Repeat;
CC Immunoglobulin domain; Proto-oncogene.
CC SIGNAL 1 18 POTENTIAL.
CC CHAIN 19 994 PROTO-ONCOGENE TYROSINE-PROTEIN KINASE
CC
CC DOMAIN 19 497 MER.
CC TRANSMEM 498 518 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 519 994 POTENTIAL.
CC DOMAIN 75 181 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 192 268 IG-LIKE C2-TYPE 1.
CC DOMAIN 279 363 IG-LIKE C2-TYPE 2.
CC DOMAIN 378 468 FIBRONECTIN TYPE-III 1.
CC DOMAIN 582 852 FIBRONECTIN TYPE-III 2.
CC NP_BIND 588 596 PROTEIN KINASE.
CC BINDING 610 610 ATP (BY SIMILARITY).
CC ACT_SITE 718 718 ATP (BY SIMILARITY).
CC DISULFID 109 170 BY SIMILARITY.
CC DISULFID 213 257 BY SIMILARITY.
CC CARBOHYD 91 91 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 108 108 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 165 165 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 202 202 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 210 210 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 229 229 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 289 289 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 311 311 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 324 324 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 331 331 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 349 349 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 384 384 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 390 390 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 437 437 N-LINKED (GLCNAC... ) (POTENTIAL).
CC CARBOHYD 449 449 N-LINKED (GLCNAC... ) (POTENTIAL).
CC MOD_RES 749 749 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```





```
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RT Nature 417:141-147(2002).
CC -!- FUNCTION: Excises uracil residues from the DNA which can arise as
CC a result of misincorporation of dUMP residues by DNA polymerase or
CC due to deamination of cytosine (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the uracil-DNA glycosylase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL939107; CAC13066.1; -.
DR HSSP; P13051; 1UGH.
DR HAMAP; MF_00148; -. 1.
DR InterPro; IPR003249; U_glycylse_notp.
DR InterPro; IPR002043; UDNA_glycylse.
DR InterPro; IPR005122; UDNA_glycylseSF.
DR Pfam; PF03167; UDG; 1.
DR ProDom; PD001589; U_glycylse_notp; 1.
DR TIGRFAMs; TIGR00628; ung; 1.
DR PROSITE; PS00130; U DNA GLYCOSYLASE; 1.
KW DNA repair; Hydrolase; Glycosylase; Complete proteome.
FT ACT SITE 68 68 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 225 AA; 24456 MW; B370AE819B12D144 CRC64;

Query Match 49.4%; Score 41; DB 1; Length 225;
Best Local Similarity 70.0%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 APDTRPAPGS 14
Db 84 AFEVRLPGS 93

RESULT 23
TRZA RHOCG
ID TRZA RHOCG STANDARD; PRT; 476 AA.
AC Q52725;
DT 15-JUL-1998 (Rel. 36, Created);
DT 15-JUL-1998 (Rel. 36, Last sequence update);
DT 28-FEB-2003 (Rel. 41, Last annotation update);
DE S-triazine hydrolase {EC 3.8.1.-} (N-ethylamine chloride hydrolase).
GN TRZA.
OS Rhodococcus corallinus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Gordoniaceae; Gordonia.
OX NCBI_TaxID=36822;
RN [1]
RP SEQUENCE FROM N.A.; AND PARTIAL SEQUENCE.
RC STRAIN=NRRL 15444R;
RX MEDLINE=96011356; PubMed=7592318;
RA Shao Z.Q., Seffens W., Mulbry W., Behki R.M.;
RT "Cloning and expression of the s-triazine hydrolase gene (trza) from
RT Rhodococcus corallinus and development of Rhodococcus recombinant
RT strains capable of dealkylating and dechlorinating the herbicide
RT atrazine.";
RL J. Bacteriol. 177:5748-5755(1995).
CC -!- FUNCTION: HYDROLYTIC DEAMINATION OF THE S-TRIAZINE SUBSTRATE
CC MELAMINE.
CC -!- PATHWAY: Melamine degradation pathway; first step.
CC -!- SIMILARITY: Belongs to the ATZ/TRZ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16534; AAA90931.1; -.
DR PIR; T46666; T46666.
DR InterPro; IPR006680; Amidohydro_1.
DR Pfam; PF01979; Amidohydro_1; 1.
KW Hydrolase.
FT INIT MET 0 0
SQ SEQUENCE 476 AA; 50727 MW; 64D953DB2E92C73E CRC64;

Query Match 49.4%; Score 41; DB 1; Length 476;
Best Local Similarity 72.7%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TSAPLTPAPG 13
Db 126 TSDPTTSPAG 136

RESULT 24
ECM1 MOUSE
ID ECM1 MOUSE STANDARD; PRT; 559 AA.
AC Q61508;
DT 01-NOV-1997 (Rel. 35, Created);
DT 01-NOV-1997 (Rel. 35, Last sequence update);
DT 28-FEB-2003 (Rel. 41, Last annotation update);
DE Extracellular matrix protein 1 precursor (Secretory component p85).
GN ECM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT), AND SEQUENCE OF 20-37;
RC STRAIN=BALB/C;
RX MEDLINE=95332352; PubMed=7608209;
RA Bhale Rao J., Tylzanowski P., Filie J.D., Kozak C.A., Merregaert J.;
RT "Molecular cloning, characterization, and genetic mapping of the cDNA
RT coding for a novel secretory protein of mouse. Demonstration of
RT alternative splicing in skin and cartilage.";
RL J. Biol. Chem. 270:16385-16394(1995).
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q61508-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q61508-2; Sequence=VSP_004230;
CC -!- TISSUE SPECIFICITY: The long isoform is expressed in a number of
CC tissues including liver, heart and lungs. The short isoform is
CC expressed in skin and cartilage-containing tissues such as tail
CC and front paw. No expression is found in brain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L33416; AAA37535.1; -.
DR MGD; MGI:103060; Ecml.
DR GO; GO:0005615; C:extracellular space; IDA.
DR InterPro; IPR008605; ECM1.
DR InterPro; IPR000264; Serum_albumin.
DR Pfam; PF05782; ECM1; 1.
KW Signal; Alternative splicing; Extracellular matrix; Glycoprotein;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 559 EXTRACELLULAR MATRIX PROTEIN 1.
```

```
FT DOMAIN 170 424 2 X APPROXIMATE REPEATS.
FT REPEAT 170 298 1.
FT REPEAT 302 424 2.
FT CARBOHYD 373 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 256 380 Missing (in isoform Short).
FT SEQUENCE 559 AA; 62775 MW; BBF37FAB7D67E2E8 CRC64;

Query Match 49.4%; Score 41; DB 1; Length 559;
Best Local Similarity 57.1%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GVTAPDTRPAPGS 14
Eb 543 GPTRGTDANPAPGS 556

RESULT 25
GUND CELFI
ID GUND_CELFI STANDARD; PRT; 747 AA.
AC P50400;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CEND.
OS Cellulomonas fimi.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93209933; PubMed=8458833;
RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.;
RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cend), a family A beta-1,4-glucanase.";
RL J. Bacteriol. 175:1910-1918(1993).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PATHWAY: Cellulose degradation.
CC -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD) domain.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl hydrolases).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
```

```
FT DOMAIN 448 542 FIBRONECTIN TYPE-III 1.
FT DOMAIN 546 639 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 208 208 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 349 349 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 747 AA; 78936 MW; BD15473C9D8B42BD CRC64;

Query Match 49.4%; Score 41; DB 1; Length 747;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 10; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 1 GVTAPDTRPAPGSTA 16
Db 633 GVTAPD--PTTGSCA 646

RESULT 26
APMU PIG
ID APMU_PIG STANDARD; PRT; 1150 AA.
AC P12021;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apomucin (Mucin core protein) (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=91236743; PubMed=2033060;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT "Porcine submaxillary mucin contains a cysteine-rich, carboxyl-terminal domain in addition to a highly repetitive, glycosylated domain.";
RL J. Biol. Chem. 266:9678-9686(1991).
RN [2]
RP SEQUENCE OF 1-503 FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=88087170; PubMed=2826455;
RA Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT "Porcine submaxillary gland apomucin contains tandemly repeated, identical sequences of 81 residues.";
RL J. Biol. Chem. 263:1081-1088(1988).
RN [3]
RP SEQUENCE OF 45-80.
RC TISSUE=Submaxillary gland;
RX MEDLINE=87280230; PubMed=3611111;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Toumadje A., Johnson W.C. Jr., Hill R.L.;
RT "Structural properties of porcine submaxillary gland apomucin.";
RL J. Biol. Chem. 262:11339-11344(1987).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
RC TISSUE=Submaxillary gland;
RX MEDLINE=97248516; PubMed=9092502;
RA Gerken T.A., Owens C.L., Pasumathy M.;
RT "Determination of the site-specific O-glycosylation pattern of the porcine submaxillary mucin tandem repeat glycopeptide. Model proposed for the polypeptide:galnac transferase peptide binding site.";
RL J. Biol. Chem. 272:9709-9719(1997).
CC -!- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL ENVIRONMENT.
CC -!- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A MULTIMERIC MUCIN STRUCTURE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -!- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81 RESIDUES.
CC -!- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
```





```
CC CC -!- COFACTOR: Calcium (By similarity).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC   Event=Alternative splicing; Named isoforms=3;
CC CC   Name=A; Synonyms=C-beta-1a;
CC CC   IsoId=Q9Z1B3-1; Sequence=Displayed;
CC CC   Name=B; Synonyms=C-beta-1b;
CC CC   IsoId=Q9Z1B3-2; Sequence=VSP_008917;
CC CC   Name=C;
CC CC   IsoId=Q9Z1B3-3; Sequence=VSP_008918;
CC CC   Note=No experimental confirmation available;
CC CC -!- MISCELLANEOUS: THE RECEPTOR-MEDIATED ACTIVATION OF PLC-BETA 1 IS
CC CC   MEDIATED BY TWO G-PROTEIN ALPHA SUBUNITS, ALPHA-Q AND ALPHA-11.
CC CC -!- SIMILARITY: Contains 1 PI-PLC X-box catalytic domain.
CC CC -!- SIMILARITY: Contains 1 PI-PLC Y-box catalytic domain.
CC CC -!- SIMILARITY: Contains 1 C2 domain.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U85712; AAD00571.1; -.
CC CC EMBL; U85713; AAD00572.1; -.
CC CC EMBL; U85714; AAD00573.1; -.
CC CC EMBL; AF498249; AAM22966.1; -.
CC CC EMBL; AF498250; AAM22967.1; -.
CC CC EMBL; AF022801; AAD01749.1; -.
CC CC EMBL; X95344; CAA64637.1; -.
CC CC PIR; S68256; S68256.
CC CC HSSP; P10688; 1DJX.
CC CC MGD; MGI:97613; P1cbl.
CC CC InterPro; IPR000008; C2.
CC CC InterPro; IPR008973; C2 CaLB.
CC CC InterPro; IPR001192; PI-PLC.
CC CC InterPro; IPR00909; PI-PLC Xdom.
CC CC InterPro; IPR001711; PI-PLC_Y.
CC CC Pfam; PF00168; C2; 1.
CC CC Pfam; PF00388; PI-PLC-X; 1.
CC CC Pfam; PF00387; PI-PLC-Y; 1.
CC CC PRINTS; PR00390; PHPLIPASEC.
CC CC ProDom; PD001202; PI-PLC_Y; 1.
CC CC SMART; SM00239; C2; 1.
CC CC SMART; SM00148; PLCXC; 1.
CC CC SMART; SM00149; PLCYC; 1.
CC CC PROSITE; PS50004; C2 DOMAIN 2; 1.
CC CC PROSITE; PS50007; PIPLC_X DOMAIN; 1.
CC CC PROSITE; PS50008; PIPLC_Y DOMAIN; 1.
CC CC Lipid degradation; Hydrolase; Transducer; Calcium; Phosphorylation;
CC CC Alternative splicing.
CC CC DOMAIN 316 467 PI-PLC X-BOX.
CC CC DOMAIN 540 656 PI-PLC Y-BOX.
CC CC DOMAIN 663 761 C2 DOMAIN.
CC CC DOMAIN 914 1088 LYS-RICH.
CC CC ACT_SITE 331 331 BY SIMILARITY.
CC CC ACT_SITE 378 378 BY SIMILARITY.
CC CC MOD_RES 887 887 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
CC CC VARSPPLIC 1142 1216 LQTELEQYQDKFKRLPLELLEFVQEAAMKGIKISDSNHGSA
CC CC PPSLASDAAKVNLKSPSEIERENPGREFDTPL -> GEG
CC CC PSSVLSEGHEDPSVPPNTPNPQALKW (in
CC CC isoform B).
CC CC /FTId=VSP_008917.
CC CC Missing (in isoform C).
CC CC /FTId=VSP_008918.
CC CC VARSPLIC 1199 1216
CC CC VARIANT 28 28 L -> F (in strains ILS and ISS).
CC CC VARIANT 41 41 T -> I (in strains C57BL/6, ILS and ISS).
CC CC VARIANT 67 67 S -> T (in strain C57BL/6).
CC CC VARIANT 79 79 E -> K (in strains C57BL/6, ILS and ISS).
CC CC VARIANT 112 112 A -> V (in strains C57BL/6, ILS and ISS).
CC CC VARIANT 622 622 M -> V (in strains ILS and ISS).
CC CC VARIANT 714 714 T -> K (in strains ILS and ISS).
```

```
FT VARIANT 795 795 D -> V (in strains ILS and ISS).
FT VARIANT 923 923 H -> Q (in strains ILS and ISS).
FT VARIANT 957 957 I -> N (in strains ILS and ISS).
FT VARIANT 1084 1084 T -> K (in strains ILS and ISS).
FT VARIANT 1197 1197 S -> G (in strain ILS).
FT VARIANT 1197 1197 S -> W (in strain ISS).
FT VARIANT 1198 1198 P -> S (in strains ILS and ISS).
FT CONFLICT 561 561 R -> I (IN REF. 4).
FT CONFLICT 613 613 L -> I (IN REF. 4).
SQ SEQUENCE 1216 AA; 138324 MW; D873078A58CE824D CRC64;

Query Match 49.4%; Score 41; DB 1; Length 1216;
Best Local Similarity 58.3%; Pred. No. 2.1e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGS 14
DB 876 SQAPHSQPAPGS 887
: |||:|||||
STANDARD; PRT; 1216 AA.

RESULT 28
PIB1_RAT
ID PIB1_RAT STANDARD; PRT; 1216 AA.
AC P10687;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta 1
DE (EC 3.1.4.11) (Phosphoinositide phospholipase C) (PLC-beta-1)
DE (Phospholipase C-beta-1) (PLC-I) (PLC-154).
GN PLCB1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88270495; PubMed=3390863;
RA Suh P.-G., Ryu S.H., Moon K.H., Suh H.W., Rhee S.G.;
RT "Cloning and sequence of multiple forms of phospholipase C.";
RL Cell 54:161-169(1988).
RN [2]
RP ACTIVITY, TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RX MEDLINE=93203266; PubMed=8454637;
RA Jhon D.-Y., Lee H.-H., Park D., Lee C.-W., Lee K.-H., Yoo O.J.,
RA Rhee S.G.;
RT "Cloning, sequencing, purification, and Gq-dependent activation of
RT phospholipase C-beta 3.";
RL J. Biol. Chem. 268:6654-6661(1993).
CC -!- FUNCTION: The production of the second messenger molecules
CC diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3) is
CC mediated by activated phosphatidylinositol-specific phospholipase
CC C enzymes.
CC -!- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-
CC bisphosphate + H(2)O = D-myo-inositol 1,4,5-trisphosphate +
CC diacylglycerol.
CC -!- COFACTOR: Calcium.
CC -!- SUBCELLULAR LOCATION: Cytosolic and particulate fractions.
CC -!- TISSUE SPECIFICITY: Highest expression in brain. Also expressed in
CC parotid gland, liver, uterus, lung, heart, adrenal gland and
CC ovary. Not detected in spleen, pancreas, intestine, thymus or
CC kidney.
CC -!- MISCELLANEOUS: THE RECEPTOR-MEDIATED ACTIVATION OF PLC-BETA 1 IS
CC MEDIATED BY TWO G-PROTEIN ALPHA SUBUNITS, ALPHA-Q AND ALPHA-11.
CC -!- SIMILARITY: Contains 1 PI-PLC X-box catalytic domain.
CC -!- SIMILARITY: Contains 1 PI-PLC Y-box catalytic domain.
CC -!- SIMILARITY: Contains 1 C2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M20636; AAA41885.1; --  
 PIR; A28821; A28821.  
 HSSP; P10688; 1DJX.  
 InterPro; IPR000008; C2.  
 InterPro; IPR008373; C2 CaLB.  
 InterPro; IPR001192; PI\_PLC.  
 InterPro; IPR000909; PI\_PLC\_Xdom.  
 InterPro; IPR001711; PI\_PLC\_Y.  
 Pfam; PF00168; C2; 1.  
 Pfam; PF00388; PI-PLC-X; 1.  
 Pfam; PF00387; PI-PLC-Y; 1.  
 PRINTS; PR00390; PHPLIPASEC.  
 ProDom; PD001202; PI\_PLC\_Y; 1.  
 SMART; SM03239; C2; 1.  
 SMART; SM00148; PLCXC; 1.  
 SMART; SM00149; PLCYC; 1.  
 PROSITE; PS50004; C2 DOMAIN 2; 1.  
 PROSITE; PS50007; PIFLC\_X\_DOMAIN; 1.  
 PROSITE; PS50008; PIFLC\_Y\_DOMAIN; 1.  
 Hydrolase; Lipid degradation; Transducer; Phosphorylation; Calcium.  
 DOMAIN 316 467  
 PI-PLC X-BOX.  
 PI-PLC Y-BOX.  
 DOMAIN 540 656  
 C2 DOMAIN.  
 ACT SITE 331 331  
 BY SIMILARITY.  
 ACT SITE 378 378  
 BY SIMILARITY.  
 MOD\_RES 887 887  
 PHOSPHORYLATION (BY PKC).  
 SEQUENCE 1216 AA; 138344 MW; 92F23691781F788E CRC64;

Query Match 49.4%; Score 41; DB 1; Length 1216;  
 Best Local Similarity 58.3%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TSAPDTRPAPGS 14  
 Db 876 SQAPHSQPAPGS 887

RESULT 29  
 AIM1\_HUMAN  
 ID AIM1\_HUMAN STANDARD; PRT; 1723 AA.  
 AC Q9Y4K1; C00296;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Absent in melanoma 1 protein.  
 GN AIM1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=97250519; PubMed=9096375;  
 RA Ray M.E., Wistow G., Su Y.A., Meltzer P.S., Trent J.M.;  
 RT "AIM1, a novel non-lens member of the betagamma-crystallin  
 RT superfamily, is associated with the control of tumorigenicity in human  
 RT malignant melanoma."  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234 (1997).  
 CC -!- FUNCTION: May function as suppressor of malignant melanoma. It may  
 CC exert its effects through interactions with the cytoskeleton.  
 CC -!- SIMILARITY: Belongs to the beta/gamma-crystallin family.  
 CC -!- SIMILARITY: Contains 12 beta/gamma crystallin 'Greek key' domains.  
 CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U83116; AAB53792.1; --  
 EMBL; U83115; AAB53791.1; --  
 HSSP; P02526; IGAM.  
 Genew; HGNC:356; AIM1.  
 MIM; 601797; --  
 InterPro; IPR001064; Crystallin.  
 InterPro; IPR000772; Ricin B lectin.  
 InterPro; IPR008997; RicinB\_like.  
 Pfam; PF00030; crystall; 6.  
 Pfam; PF00652; Ricin B lectin; 3.  
 PRINTS; PR01367; BGCRCYSTALLIN.  
 SMART; SM00458; RICIN; 1.  
 SMART; SM00247; XTALB9; 6.  
 PROSITE; PS50915; CRYSTALLIN BETAGAMMA; 12.  
 PROSITE; PS50231; RICIN\_B\_LECTIN; 1.  
 Repeat; Lectin.  
 DOMAIN 1022 1061  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 1.  
 DOMAIN 1052 1117  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 2.  
 DOMAIN 1123 1163  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 3.  
 DOMAIN 1154 1206  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 4.  
 DOMAIN 1218 1270  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 5.  
 DOMAIN 1271 1313  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 6.  
 DOMAIN 1319 1361  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 7.  
 DOMAIN 1362 1404  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 8.  
 DOMAIN 1415 1452  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 9.  
 DOMAIN 1453 1496  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 10.  
 DOMAIN 1502 1542  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 11.  
 DOMAIN 1543 1584  
 BETA/GAMMA CRYSTALLIN 'GREEK KEY' 12.  
 RICIN B-TYPE LECTIN.  
 SEQUENCE 1723 AA; 188646 MW; 7E50F681A627FB09 CRC64;

Query Match 49.4%; Score 41; DB 1; Length 1723;  
 Best Local Similarity 50.0%; Pred. No. 3.1e+02;  
 Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

Qy 1 GWTSA--PDTRPAPGS 14  
 Db 69 GVASAASPEKPSGT 84

RESULT 30  
 PGHM\_HUMAN  
 ID PGHM\_HUMAN STANDARD; PRT; 4391 AA.  
 AC P98160; Q16287; Q9H3V5;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Basement membrane-specific heparan sulfate proteoglycan core  
 DE protein precursor (HSPG) (Perlecan) (PLC).  
 GN HSPG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=92112994; PubMed=1730768;  
 RX Kallunki P., Tryggvason K.;  
 RT "Human basement membrane heparan sulfate proteoglycan core protein: a  
 RT 467-kD protein containing multiple domains resembling elements of the  
 RT low density lipoprotein receptor, laminin, neural cell adhesion  
 RT molecules, and epidermal growth factor."  
 RL J. Cell Biol. 116:559-571 (1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon, and Skin;  
 RX MEDLINE=92235084; PubMed=1569102;  
 RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;  
 RT "Primary structure of the human heparan sulfate proteoglycan from  
 RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple

domains homologous to the low density lipoprotein receptor, laminin, neural cell adhesion molecules, and epidermal growth factor.";  
J. Biol. Chem. 267:8544-8557(1992).  
[3]  
SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.  
MEDLINE=20553141; PubMed=1101850;  
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,  
Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,  
Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,  
Hentati F., Fontaine B.;  
"Perlecan, the major proteoglycan of basement membranes, is altered in patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";  
Nat. Genet. 26:480-483(2000).  
[4]  
SEQUENCE OF 1015-1470 FROM N.A.  
MEDLINE=91365376; PubMed=1679749;  
RA Dodge G.R., Kovalsky I., Chu M.L., Hassell J.R., McBride O.W.,  
Yi H.F., Iozzo R.V.;  
"Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellular expression, and mapping of the gene (HSPG2) to the short arm of human chromosome 1.";  
Genomics 10:673-680(1991).  
[5]  
SEQUENCE OF 890-1396 FROM N.A.  
TISSUE=Fibrosarcoma;  
MEDLINE=92120660; PubMed=1685141;  
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,  
Tryggvason K.;  
"Cloning of human heparan sulfate proteoglycan core protein, assignment of the gene (HSPG2) to 1p36.1--p35 and identification of a BamHI restriction fragment length polymorphism.";  
Genomics 11:389-396(1991).  
[6]  
SEQUENCE OF 1-21 FROM N.A.  
MEDLINE=94052171; PubMed=8234307;  
RA Cohen I.R., Graessell S., Murdoch A.D., Iozzo R.V.;  
"Structural characterization of the complete human perlecan gene and its promoter.";  
Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).  
[7]  
CARBOHYDRATE-LINKAGE SITE ASN-2121.  
MEDLINE=22660472; PubMed=12754519;  
RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;  
"Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry.";  
Nat. Biotechnol. 21:660-666(2003).  
-!- FUNCTION: This protein is an integral component of basement membranes. It is responsible for the fixed negative electrostatic charge and is involved in the charge-selective ultrafiltration properties. It serves as an attachment substrate for cells.  
-!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in dimers or stellate structures. It interacts with other basement membrane components such as laminin, prolargin and collagen type IV.  
-!- SUBCELLULAR LOCATION: Extracellular.  
-!- TISSUE SPECIFICITY: Found in the basement membranes.  
-!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-LINKED OLIGOSACCHARIDES.  
-!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder characterized by permanent myotonia (prolonged failure of muscle relaxation) and skeletal dysplasia, resulting in reduced stature, kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.  
-!- SIMILARITY: Contains 4 LDL-receptor class A domains.  
-!- SIMILARITY: Contains 11 laminin EGF-like domains.  
-!- SIMILARITY: Contains 3 laminin IV domains.  
-!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.  
-!- SIMILARITY: Contains 3 laminin G-like domains.  
-!- SIMILARITY: Contains 4 EGF-like domains.  
-!- SIMILARITY: Contains 1 SEA domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; X62515; CAA44373.1; -.  
CC EMBL; M85289; AAA52700.1; -.  
CC EMBL; AL445795; CAC18534.1; -.  
CC EMBL; M64283; AAA52699.1; -.  
CC EMBL; S76436; AAB21121.2; -.  
CC EMBL; L22078; -; NOT\_ANNOTATED\_CDS.  
CC PIR; A38096; A38096.  
CC HSSP; P00740; 1EDM.  
CC Siena-2DPAGE; F98160; -.  
CC Genew; HGNC:5273; HSPG2.  
CC MIM; 142461; -.  
CC MIM; 255800; -.  
CC InterPro; IPR008985; ConA\_like\_lec\_gl.  
CC InterPro; IPR00742; EGF\_2.  
CC InterPro; IPR006209; EGF\_like.  
CC InterPro; IPR006210; IEGF.  
CC InterPro; IPR007110; Ig-like.  
CC InterPro; IPR003599; Ig.  
CC InterPro; IPR003598; Ig\_c2.  
CC InterPro; IPR003596; Ig\_v.  
CC InterPro; IPR000034; Laminin\_B.  
CC InterPro; IPR002049; Laminin\_EGF.  
CC InterPro; IPR001791; Laminin\_G.  
CC InterPro; IPR002172; LDL\_receptor\_A.  
CC InterPro; IPR000082; SEA\_domain.  
CC Pfam; PF00008; EGF; 4.  
CC Pfam; PF00047; Ig; 22.  
CC Pfam; PF00052; laminin\_B; 3.  
CC Pfam; PF00053; laminin\_EGF; 7.  
CC Pfam; PF00054; laminin\_G; 3.  
CC Pfam; PF00057; ldl\_recept\_a; 4.  
CC Pfam; PF01390; SEA; 1.  
CC PRINTS; PR00261; LDLRECEPTOR.  
CC ProDom; PD003031; Laminin\_B; 3.  
CC SMART; SM00181; EGF; 15.  
CC SMART; SM00180; EGF\_Lam; 12.  
CC SMART; SM00409; IG; 22.  
CC SMART; SM00408; IGC2; 21.  
CC SMART; SM00406; IGV; 7.  
CC SMART; SM00281; LamB; 3.  
CC SMART; SM00282; LamG; 3.  
CC SMART; SM00192; LDLa; 4.  
CC SMART; SM00200; SEA; 1.  
CC PROSITE; PS00022; EGF\_1; 9.  
CC PROSITE; PS01186; EGF\_2; 6.  
CC PROSITE; PS00026; EGF\_3; 4.  
CC PROSITE; PS00835; IG LIKE; 22.  
CC PROSITE; PS00025; LAM\_G\_DOMAIN; 3.  
CC PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.  
CC PROSITE; PS01209; LDLRA\_1; 4.  
CC PROSITE; PS00068; LDLRA\_2; 4.  
CC PROSITE; PS00024; SEA; 1.  
KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;  
KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;  
KW Extracellular matrix; EGF-like domain; Disease mutation.  
FT SIGNAL 1 21 POTENTIAL  
FT CHAIN 22 4391 BASEMENT MEMBRANE-SPECIFIC HEPARAN  
FT DOMAIN 80 194 SULFATE PROTEOGLYCAN CORE PROTEIN.  
FT DOMAIN 198 235 SEA.  
FT DOMAIN 284 320 LDL-RECEPTOR CLASS A 1.  
FT DOMAIN 324 360 LDL-RECEPTOR CLASS A 2.  
FT DOMAIN 367 404 LDL-RECEPTOR CLASS A 3.  
FT DOMAIN 405 504 LDL-RECEPTOR CLASS A 4.  
FT DOMAIN 521 530 IG-LIKE C2-TYPE 1.  
FT DOMAIN 531 730 LAMININ EGF-LIKE 1 (N-TERMINAL).  
FT DOMAIN LAMININ DOMAIN IV 1 (DOMAIN III A).



RA	Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M., Sasaki T., Tajima N., Iwanaga T., Seino S., Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of CAMP-GEFII.Rim2.Piccolo complex in cAMP-dependent exocytosis.";	CC	J. Biol. Chem. 277:50497-50502(2002).	CC	-!- FUNCTION: May act as a scaffolding protein involved in the organization of synaptic active zones and in synaptic vesicle trafficking.	CC	-!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin.	CC	-!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of synaptic junctions.	CC	-!- ALTERNATIVE PRODUCTS:	CC	Event=Alternative splicing; Named isoforms=2;	CC	Name=1;	CC	IsoId=Q9QYX7-1; Sequence=Displayed;	CC	Name=2;	CC	IsoId=Q9QYX7-2; Sequence=VSP_003928, VSP_003929;	CC	-!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in stomach. Not detected in other tissues analyzed including adrenal gland, testis and pancreas.	CC	-!- DOMAIN: C2 domain 1 is involved in binding calcium and phospholipids. Calcium binds with low affinity but with high specificity and induces a large conformational change.	CC	-!- SIMILARITY: Contains 2 C2 domains.	CC	-!- SIMILARITY: Contains 1 PDZ/DHR domain.	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	CC	EMBL; Y19185; CAB60731.2; --	DR	EMBL; Y19186; CAB60732.2; --	DR	EMBL; AF181269; AAD55786.2; --	DR	HSSP; P04410; 1A25.	DR	MGD; MGI:1349390; Pclo.	DR	GO; GO:0045202; C:synaptic junction; IDA.	DR	GO; GO:0005509; F:calcium ion binding; ISS.	DR	GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.	DR	GO; GO:0005522; F:profilin binding; IDA.	DR	GO; GO:0019933; P:cAMP-mediated signaling; IDA.	DR	GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.	DR	GO; GO:0030073; P:insulin secretion; IDA.	DR	GO; GO:0017157; P:regulation of exocytosis; IDA.	DR	GO; GO:0016080; P:synaptic vesicle targeting; NAS.	DR	InterPro; IPR000008; C2.	DR	InterPro; IPR001478; PDZ.	DR	InterPro; IPR008899; Znf_piccolo.	DR	Pfam; PF00168; C2; 2.	DR	Pfam; PF00595; PDZ; 1.	DR	Pfam; PF05715; ZF_piccolo; 2.	DR	SMART; SM00239; C2; 2.	DR	SMART; SM00228; PDZ; 1.	DR	PROSITE; PS00499; C2_DOMAIN_1; 1.	DR	PROSITE; PS50004; C2_DOMAIN_2; 2.	DR	PROSITE; PS50106; PDZ; 1.	KW	Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger; Repeat; Alternative splicing.	KW	DOMAIN 371 470 10 X 10 AA TANDEM APPROXIMATE REPEATS OF P-A-K-P-Q-P-Q-P-X.	FT	DOMAIN 371 470 10 X 10 AA TANDEM APPROXIMATE REPEATS OF P-A-K-P-Q-P-Q-P-X.	FT	ZN_FING 502 526 C4-TYPE (POTENTIAL).	FT	ZN_FING 967 990 C4-TYPE (POTENTIAL).	FT	DOMAIN 2305 2329 POLY-PRO.	FT	DOMAIN 4394 4488 PDZ.	FT	DOMAIN 4607 4705 C2_DOMAIN 1.	FT	DOMAIN 4922 5012 C2_DOMAIN 2.	FT	VARSPIC 4829 4833 TKPTN -> SKRRK (in isoform 2).	FT	VARSPIC 4834 5038 /FTId=VSP_003928.	FT	VARSPIC 4834 5038 Missing (in isoform 2).	FT	VARSPIC 4834 5038 /FTId=VSP_003929.
----	--	----	---------------------------------------	----	---	----	--	----	---	----	---------------------------	----	---	----	---------	----	-------------------------------------	----	---------	----	--	----	--	----	--	----	--	----	--	----	--	----	------------------------------	----	------------------------------	----	--------------------------------	----	---------------------	----	-------------------------	----	---	----	---	----	--	----	--	----	---	----	--	----	---	----	--	----	--	----	--------------------------	----	---------------------------	----	-----------------------------------	----	-----------------------	----	------------------------	----	-------------------------------	----	------------------------	----	-------------------------	----	-----------------------------------	----	-----------------------------------	----	---------------------------	----	---	----	--	----	--	----	--------------------------------------	----	--------------------------------------	----	----------------------------	----	-----------------------	----	-------------------------------	----	-------------------------------	----	--	----	-------------------------------------	----	---	----	-------------------------------------

SQ SEQUENCE 5038 AA; 547600 MW; DADA460CF3B40888 CRC64;  
Query Match 49.4%; Score 41; DB 1; Length 5038;  
Best Local Similarity 63.6%; Pred. No. 9.4e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 VTSAPDTRPAP 12  
|||:|  
Db 770 VTPKPDTPKVP 780

RESULT 32  
CLR3 MOUSE  
ID CLR3 MOUSE STANDARD; PRT; 3301 AA.  
AC Q91ZIO; Q9ESD0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor.  
GN CELSR3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGES.  
RC STRAIN=C57BL/6;  
RX MEDLINE=21839555; PubMed=11850187;  
RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;  
RT "Developmental expression profiles of Celsr (Flamingo) genes in the mouse."  
RL Mech. Dev. 112:157-160(2002).  
RN [2]  
RP SEQUENCE OF 2574-3046 FROM N.A., AND DEVELOPMENTAL STAGE.  
RX MEDLINE=21534880; PubMed=11677057;  
RA Formstone C.J., Little P.F.R.;  
RT "The flamingo-related mouse Celsr family (Celsr1-3) genes exhibit distinct patterns of expression during embryonic development."  
RL Mech. Dev. 109:91-94(2001).  
RN [3]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=20253755; PubMed=10790539;  
RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;  
RT "Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is a candidate for the tippy (tip) lethal mutant on chromosome 9."  
RL Mamm. Genome 11:392-394(2000).  
CC -!- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.  
CC -!- DEVELOPMENTAL STAGE: Predominantly expressed in the CNS, the emerging dorsal root ganglia and cranial ganglia. In the CNS, expression is uniform along the rostrocaudal axis. No expression is detected until somite stages. Between E10 and E12, expression is strong in the marginal zone (MZ), and lower in the ventricular zone (VZ). At E15, expression is restricted to the brain and olfactory epithelium. In the brain, it is low in VZ but strong in external fields, particularly those with ongoing migration, such as the telencephalic cortical plate, the olfactory bulb, the cerebellum and the tectum. In the newborn and postnatal stages, expression is high in differentiated neuronal fields.  
CC -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.  
CC -!- SIMILARITY: Contains 9 cadherin domains.  
CC -!- SIMILARITY: Contains 7 EGF-like domains.  
CC -!- SIMILARITY: Contains 2 laminin G-like domains.  
CC -!- SIMILARITY: Contains 1 laminin EGF-like domain.  
CC -!- SIMILARITY: Contains 1 GPS domain.  
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts in positions 2575 and 2578.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF427498; AAL25099.1; -. ALT\_FRAME.  
DR EMBL; AF188752; AAG17057.1; Celser3.  
DR MGI; MGI:1858236; Celser3.  
DR InterPro; IPR000152; Asx\_hydroxyl\_s.  
DR InterPro; IPR002126; Cadherin.  
DR InterPro; IPR008985; ConA\_like\_lect\_gl.  
DR InterPro; IPR000742; EGF\_2.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR000832; GPCR\_secretin.  
DR InterPro; IPR001879; horan\_receptor.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR InterPro; IPR001791; Laminin\_G.  
DR InterPro; IPR000203; PKD\_cys\_rich.  
DR Pfam; PF00002; 7tm\_2; 1.  
DR Pfam; PF00028; cadherin; 9.  
DR Pfam; PF00008; EGF; 5.  
DR Pfam; PF01825; GPS; 1.  
DR Pfam; PF02793; HRM; 1.  
DR Pfam; PF00053; laminin\_EGF; 1.  
DR Pfam; PF00054; laminin\_G; 2.  
DR PRINTS; PR00205; CADHERIN.  
DR PRINTS; PR00249; GPCRSECRETIN.  
DR SMART; SM00112; CA; 9.  
DR SMART; SM00181; EGF; 6.  
DR SMART; SM00303; GPS; 1.  
DR SMART; SM00008; HormR; 1.  
DR SMART; SM00282; LamG; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00232; CADHERIN\_1; 7.  
DR PROSITE; PS00268; CADHERIN\_2; 8.  
DR PROSITE; PS00032; EGF\_1; 5.  
DR PROSITE; PS01186; EGF\_2; 4.  
DR PROSITE; PS00036; EGF\_3; 6.  
DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.  
DR PROSITE; PS00227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
DR PROSITE; PS00261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
DR PROSITE; PS00231; GPS; 1.  
DR PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
KW Developmental protein; Hydroxylation; Signal.  
FT SIGNAL 1 31 POTENTIAL.  
FT CHAIN 32 3301 CADHERIN EGF LAG SEVEN-PASS G-TYPE RECEPTOR 3.  
FT DOMAIN 32 2531 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 2532 2552 1 (POTENTIAL).  
FT DOMAIN 2553 2563 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 2564 2584 2 (POTENTIAL).  
FT DOMAIN 2585 2592 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 2593 2613 3 (POTENTIAL).  
FT DOMAIN 2614 2634 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 2635 2655 4 (POTENTIAL).  
FT DOMAIN 2656 2673 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 2674 2694 5 (POTENTIAL).  
FT DOMAIN 2695 2716 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 2717 2737 6 (POTENTIAL).  
FT DOMAIN 2738 2744 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 2745 2765 7 (POTENTIAL).  
FT DOMAIN 2766 3301 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 317 424 CADHERIN 1.  
FT DOMAIN 425 536 CADHERIN 2.  
FT DOMAIN 537 642 CADHERIN 3.  
FT DOMAIN 643 747 CADHERIN 4.  
FT DOMAIN 748 849 CADHERIN 5.  
FT DOMAIN 850 952 CADHERIN 6.



```
FT DOMAIN 953 1058 CADHERIN 7.
FT DOMAIN 1059 1160 CADHERIN 8.
FT DOMAIN 1161 1257 CADHERIN 9.
FT DOMAIN 1366 1424 EGF-LIKE 1, CALCIUM-BINDING.
FT DOMAIN 1426 1460 EGF-LIKE 2, CALCIUM-BINDING.
FT DOMAIN 1464 1503 EGF-LIKE 3, CALCIUM-BINDING.
FT DOMAIN 1504 1708 LAMININ G-LIKE 1.
FT DOMAIN 1711 1747 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 1751 1933 LAMININ G-LIKE 2.
FT DOMAIN 1935 1971 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 1972 2002 EGF-LIKE 6, CALCIUM-BINDING.
FT DOMAIN 2003 2042 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 2044 2079 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 2085 2118 LAMININ EGF-LIKE.
FT DOMAIN 2468 2520 GPS.
FT DOMAIN 2720 2724 POLY-LEU.
FT DISULFID 1370 1381 BY SIMILARITY.
FT DISULFID 1375 1412 BY SIMILARITY.
FT DISULFID 1414 1423 BY SIMILARITY.
FT DISULFID 1430 1441 BY SIMILARITY.
FT DISULFID 1435 1450 BY SIMILARITY.
FT DISULFID 1452 1459 BY SIMILARITY.
FT DISULFID 1468 1479 BY SIMILARITY.
FT DISULFID 1473 1489 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1715 1726 BY SIMILARITY.
FT DISULFID 1720 1735 BY SIMILARITY.
FT DISULFID 1737 1746 BY SIMILARITY.
FT DISULFID 1939 1950 BY SIMILARITY.
FT DISULFID 1944 1959 BY SIMILARITY.
FT DISULFID 1961 1970 BY SIMILARITY.
FT DISULFID 1974 1985 BY SIMILARITY.
FT DISULFID 1979 1997 BY SIMILARITY.
FT DISULFID 1999 2008 BY SIMILARITY.
FT DISULFID 2016 2029 BY SIMILARITY.
FT DISULFID 2031 2041 BY SIMILARITY.
FT DISULFID 2048 2063 BY SIMILARITY.
FT DISULFID 2050 2066 BY SIMILARITY.
FT DISULFID 2068 2078 BY SIMILARITY.
FT MOD_RES 1952 1952 HYDROXYLATION (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 838 838 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1702 1702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1759 1759 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2042 2042 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2166 2166 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2185 2185 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2375 2375 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2465 2465 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2497 2497 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 2713 2713 L -> LR (IN REF. 2).
FT CONFLICT 3024 3024 R -> P (IN REF. 2).
SQ SEQUENCE 3301 AA; 358455 MW; A6B18F2DF7F4DEB6 CRC64;

Query Match 48.8%; Score 40.5; DB 1; Length 3301;
Best Local Similarity 64.3%; Pred. No. 7.2e+02;
Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 1 GVTSAP-DTRPAPG 13
D5 2779 GKXAPEETRAPG 2792

RESULT 33
CLR3 RAT
ID CLR3 RAT STANDARD; PRT; 3313 AA.
AC C88278;
DT 28-FEB-2003 (Rel. 41, Created)
```

28-FEB-2003 (Rel. 41, last sequence update)  
10-OCT-2003 (Rel. 42, last annotation update)  
Cadherin EGF IAG seven-pass G-type receptor 3 precursor {Multiple  
epidermal growth factor-like domains 2}.  
CELSR3 OR MEGF2.  
Rattus norvegicus (Rat).  
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
NCBI\_TaxID=10116;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=Sprague-Dawley; TISSUE=Brain;  
MEDLINE=98360089; PubMed=9693030;  
Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;  
"Identification of high-molecular-weight proteins with multiple  
EGF-like motifs by motif-trap screening.";  
Genomics 51:27-34(1998).  
-!- FUNCTION: Receptor that may have an important role in cell/cell  
signaling during nervous system formation.  
-!- SUBCELLULAR LOCATION: Integral membrane protein.  
-!- TISSUE SPECIFICITY: Expressed in the brain. Expressed in  
cerebellum, olfactory bulb, cerebral cortex, hippocampus and  
brain stem.  
-!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.  
-!- SIMILARITY: Contains 9 cadherin domains.  
-!- SIMILARITY: Contains 8 EGF-like domains.  
-!- SIMILARITY: Contains 2 laminin G-like domains.  
-!- SIMILARITY: Contains 1 laminin EGF-like domain.  
-!- SIMILARITY: Contains 1 GPS domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AB011528; BAA32459.1; --  
HSSP; P00740; 1EDM.  
InterPro; IPR000152; Asx\_hydroxyl\_s.  
InterPro; IPR002126; Cadherin.  
InterPro; IPR008985; ConA\_like\_lec\_gl.  
InterPro; IPR00742; EGF\_2.  
InterPro; IPR01881; EGF\_Ca.  
InterPro; IPR006209; EGF\_like.  
InterPro; IPR00832; GPCR\_secretin.  
InterPro; IPR01879; hornu\_receptor.  
InterPro; IPR06210; IEGF.  
InterPro; IPR002049; Laminin\_EGF.  
InterPro; IPR01791; Laminin\_G.  
InterPro; IPR000203; PKD\_cys\_rich.  
Pfam; PF00002; 7tm\_2; 1.  
Pfam; PF00028; cadherin; 9.  
Pfam; PF00008; EGF; 5.  
Pfam; PF01825; GPS; 1.  
Pfam; PF02793; HRM; 1.  
Pfam; PF00054; laminin\_G; 1.  
PRINTS; PR00205; CADHERIN.  
PRINTS; PR00011; EGFLAMININ.  
PRINTS; PR00249; GPCRSECRETIN.  
SMART; SM00112; CA; 9.  
SMART; SM00181; EGF; 6.  
SMART; SM00303; GPS; 1.  
SMART; SM00008; HornuR; 1.  
SMART; SM00282; LamG; 2.  
PROSITE; PS00010; ASX\_HYDROXYL; 1.  
PROSITE; PS00232; CADHERIN\_1; 7.  
PROSITE; PS0268; CADHERIN\_2; 8.  
PROSITE; PS00022; EGF\_1; 6.  
PROSITE; PS0186; EGF\_2; 4.  
PROSITE; PS00026; EGF\_3; 6.  
PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.



RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
RA Haberman K., Murray J., Johnson D., Rohlfing T., Nelson J.,  
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,  
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,  
RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,  
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,  
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,  
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,  
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,  
RA Ransperger U., Wedler H., Balke K., Wedler E., Peters S.,  
RA van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,  
RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,  
RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,  
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.,  
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
RT thaliana.";  
RT Nature 408:823-826(2000).  
RN [5]  
RC SEQUENCE FROM N.A.  
RX STRAIN=cv. Columbia;  
RX MEDLINE=22954850; PubMed=14593172;  
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
RT "Empirical analysis of transcriptional activity in the Arabidopsis  
RT genome.";  
RT Science 302:842-846(2003).  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.A.;  
RT "Full-length cDNA from Arabidopsis thaliana.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP GPI-ANCHOR.  
RX MEDLINE=22690167; PubMed=12805588;  
RA Borner G.H., Lilley K.S., Stevens T.J., Dupree P.;  
RT "Identification of glycosylphosphatidylinositol-anchored proteins in  
RT Arabidopsis. A proteomic and genomic analysis.";  
RL Plant Physiol. 132:568-577(2003).  
CC -!- FUNCTION: Probably acts as an electron carrier.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -!- DEVELOPMENTAL STAGE: Maximum levels are found in 35 day old  
CC plantlets when the rosette is mature, consisting of 8-10 fully  
CC expanded leaves, and as the floral stem starts to form. This level  
CC remains constant during the further life span of the plant.  
CC -!- INDUCTION: By dark adaptation. This gives a 20-fold increase in  
CC expression.  
CC -!- SIMILARITY: Contains 1 plastocyanin-like domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
EMBL; Z15058; CAA78771.1; --

DR EMBL; Y18227; CAA77089.1; --  
DR EMBL; AB035137; BAA86999.1; --  
DR EMBL; AF296825; -; NOT ANNOTATED\_CDS.  
DR EMBL; AY052681; AAK96585.1; --  
DR EMBL; AY034986; AAK59491.1; --  
DR EMBL; AY142577; AAN13146.1; --  
DR EMBL; AY088549; AAM66081.1; --  
DR PIR; I39698; --39698.  
DR PIR; T51838; T51838.  
DR HSSP; P29602; IJER.  
DR InterPro; IPR00923; BlueCu\_1.  
DR InterPro; IPR008972; Cupredoxin.  
DR InterPro; IPR003245; Plcyanin-like.  
DR Pfam; PF02298; Cu\_bind\_like; 1.  
DR ProDom; PD003122; Plcyanin\_like; 1.  
DR PROSITE; PS00196; COPPER\_BLUE; 1.  
KW Electron transport; Membrane; Metal-binding; Copper; Signal;  
KW Glycoprotein; GPI-anchor; Lipoprotein.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 174 BLUE COPPER PROTEIN.  
FT PROPEP 175 196 REMOVED IN MATURE FORM (PROBABLE).  
FT DOMAIN 23 118 PLASTOCYANIN-LIKE.  
FT METAL 56 66 COPPER (BY SIMILARITY).  
FT METAL 107 107 COPPER (BY SIMILARITY).  
FT METAL 112 112 COPPER (BY SIMILARITY).  
FT METAL 117 117 COPPER (BY SIMILARITY).  
FT DISULFID 79 113 BY SIMILARITY.  
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT LIPID 174 174 GPI-anchor amidated asparagine (Probable).  
FT CONFLICT 44 44 T -> S (IN REF. 1).  
FT CONFLICT 134 134 P -> L (IN REF. 1).  
FT CONFLICT 142 142 P -> L (IN REF. 1).  
SQ SEQUENCE 196 AA; 20053 MW; 05100B50518F0A56 CRC64;  
  
Query Match 48.2%; Score 40; DB 1; Length 196;  
Best Local Similarity 53.3%; Pred. No. 45;  
Matches 8; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
QY 1 GVTSPDTPRGST 15  
Db 131 GATPGAGATPARGST 145  
  
RESULT 35  
UNG\_MYCLE  
ID UNG\_MYCLE STANDARD; PRT; 227 AA.  
AC Q9CBS3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG).  
GN UNG OR ML1675.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RX MEDLINE=21126732; PubMed=11234002;  
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Easham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Fornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RL Nature 409:1007-1011(2001).  
CC -!- FUNCTION: Excises uracil residues from the DNA which can arise as

CC a result of misincorporation of dUMP residues by DNA polymerase or  
 CC due to deamination of cytosine (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the uracil-DNA glycosylase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: AL583923; CAC30628.1; -.  
 DR PIR: E87118; E87118.  
 DR KSSP: P13051; LAKZ.  
 DR Leproma; MLI675; -.  
 DR HAMAP; MF\_00148; -; 1.  
 DR InterPro; IPR003249; U\_glycylse\_notp.  
 DR InterPro; IPR002043; UDNA\_glycylse.  
 DR InterPro; IPR005122; UDNA\_glycylseSF.  
 DR Pfam; PF03167; UDG; 1.  
 DR ProDom; PD001589; U\_glycylse\_notp; 1.  
 DR TIGRFAMs; TIGR00628; ung; 1.  
 DR PROSITE; PS00130; U\_DNA\_GLYCOSYLASE; 1.  
 KW DNA repair; Hydrolase; Glycosidase; Complete proteome.  
 FT ACT SITE 68 68 GENERAL BASE (BY SIMILARITY).  
 SQ SEQUENCE 227 AA; 24655 MW; F54B7F0AE1CE43BF CRC64;

Query Match 48.2%; Score 40; DB 1; Length 227;  
 Best Local Similarity 66.7%; Pred. No. 52;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 APDTPAPGSTA 16  
 |||||  
 Db 84 APDVRPLPRSLA 95

## RESULT 36

GLTI ECOLI STANDARD; PRT; 302 AA.  
 AC P37902; P41408; P77612; Q9R7T1; Q9R7T3;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamate/aspartate periplasmic binding protein precursor.  
 GN GLTI OR B0655.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R.; Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,  
 RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
 RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,

RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 12.7-28.0 min region of the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [4]  
 RP SEQUENCE OF 242-302 FROM N.A.  
 RC STRAIN=K12 / BK9MDG;  
 RA Lum D., Wallace B.J.;

RT "Sequence and characterisation of three genes of a glutamate-aspartate  
 RT binding protein-dependent transport system of Escherichia coli K12.";  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 23-33.  
 RC STRAIN=K12 / W3110;  
 RA Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,  
 RA Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,  
 RA Hochstrasser D.F.;

RL Submitted (SEP-1994) to Swiss-Prot.  
 RN [6]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=99420855; PubMed=10493123;

RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;  
 RT "Enrichment of low abundance proteins of Escherichia coli by  
 RT hydroxyapatite chromatography.";  
 RL Electrophoresis 20:2181-2195(1999).  
 RN [7]  
 RP CHARACTERIZATION.  
 RC STRAIN=K12 / W3092;  
 RX MEDLINE=75133470; PubMed=1091635;  
 RA Willis R.C., Furlong C.E.;

RT "Purification and properties of a periplasmic glutamate-aspartate  
 RT binding protein from Escherichia coli K12 strain W3092.";  
 RL J. Biol. Chem. 250:2574-2580(1975).  
 RN [8]  
 RP CHARACTERIZATION.  
 RX MEDLINE=75133471; PubMed=1091636;

RA Willis R.C., Furlong C.E.;  
 RT "Interactions of a glutamate-aspartate binding protein with the  
 RT glutamate transport system of Escherichia coli.";  
 RL J. Biol. Chem. 250:2581-2586(1975).  
 CC -!- FUNCTION: Part of the binding-protein-dependent transport system  
 CC for glutamate and aspartate. Binds to both aspartate and  
 CC glutamate.

CC -!- SUBCELLULAR LOCATION: Periplasmic.  
 CC -!- SIMILARITY: Belongs to the bacterial extracellular solute-binding  
 CC protein family 3.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: AE000170; AAC73756.1; -.  
 DR EMBL; U82598; AAB40856.1; -.  
 DR EMBL; D90705; BAA35307.1; ALT\_INIT.  
 DR EMBL; D90706; BAA35311.1; ALT\_INIT.  
 DR EMBL; U10981; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; E64800; E64800.  
 DR SWISS-2DPAGE; P37902; COLI.  
 DR EcoGene; EG12700; glti.  
 DR InterPro; IPR001311; SBP/glu\_receptor.  
 DR InterPro; IPR001638; SBP\_bac\_3.  
 DR Pfam; PF00497; SBP\_bac\_3; 1.  
 DR SMART; SM00062; PBPb; 1.  
 DR PROSITE; PS01039; SBP\_BACTERIAL\_3; FALSE\_NEG.  
 DR Transport; Amino-acid transport; Periplasmic; Signal;  
 KW



KW Complete proteome. 22  
 FT SIGNAL 23 302 GLUTAMATE/ASPARTATE PERIPLASMIC BINDING  
 FT CHAIN 23 302 PROTEIN.  
 SQ SEQUENCE 302 AA; 33420 MW; 6FFC3321040152EC CRC64;

Query Match 48.2%; Score 40; DB 1; Length 302;  
 Best Local Similarity 53.3%; Pred. No. 71;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GVTSAPTDTRPAGST 15  
 Db 18 GLAQADDAAPAGST 32

RESULT 37  
 LIPB DEIRA  
 ID LIPB DEIRA STANDARD; PRT; 336 AA.  
 AC Q9RWAS;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lipote-protein ligase B (EC 6.-.-) (Lipoate biosynthesis protein  
 DE B).  
 GN LIPB OR DR0764.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 CC -!- FUNCTION: Involved in the attachment of lipoyl groups to proteins,  
 CC by creating an amide linkage that joins the free carboxyl group of  
 CC lipoic acid to the epsilon-amino group of a specific lysine  
 CC residue in lipoylated proteins (By similarity).  
 CC -!- PATHWAY: Lipoyate biosynthesis.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE LIPB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB001931; AAF10340.1; -.  
 CC PIR; H75479; H75479.  
 CC TIGR; DR0764; -.  
 CC HAMAP; MF 00013; fused; 1.  
 CC InterPro; IPR004143; BPL\_LipA\_LipB.  
 CC InterPro; IPR000544; Lipoate B.  
 CC Pfam; PF03099; BPL\_LipA\_LipB; 1.  
 CC ProDom; PD006086; Lipoate B; 1.  
 CC TIGRFAMs; TIGR00214; lipB; 1.  
 CC PROSITE; PS01313; LIPB; 1.  
 CC Ligase; Complete proteome.  
 KW DOMAIN 1 92 UNKNOWN.  
 FT DOMAIN 93 336 LIPB DOMAIN.  
 FT SEQUENCE 336 AA; 36178 MW; A3E454F0ACF9AA68 CRC64;

Query Match 48.2%; Score 40; DB 1; Length 336;  
 Best Local Similarity 53.8%; Pred. No. 79;  
 Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GVTSAPTDTRPAPG 13  
 Db 210 GTLGLPDARENPG 222

RESULT 38  
 FMLR MOUSE  
 ID FMLR MOUSE STANDARD; PRT; 364 AA.  
 AC P33766;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE fMet-Leu-Phe receptor (fMLP receptor) (N-formyl peptide receptor)  
 DE (FPR) (N-formylpeptide chemoattractant receptor).  
 GN FPR1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94064602; PubMed=8244972;  
 RA Gao J.-L., Murphy P.M.;  
 RT "Species and subtype variants of the N-formyl peptide chemotactic  
 RT receptor reveal multiple important functional domains.";  
 RL J. Biol. Chem. 268:25395-25401(1993).  
 CC -!- FUNCTION: High affinity receptor for N-formyl-methionyl peptides,  
 CC which are powerful neutrophils chemotactic factors. Binding of  
 CC fMLP to the receptor causes activation of neutrophils. This  
 CC response is mediated via a G-protein that activates a  
 CC phosphatidylinositol-calcium second messenger system.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L22181; AAL16110.1; -.  
 CC PIR; A49542; A49542.  
 CC MGD; MGI:107443; Fpr1.  
 CC GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IDA.  
 CC InterPro; IPR000276; GPCR\_Rhodpsn.  
 CC Pfam; PF00001; 7tm 1; 1.  
 CC PRINTS; PR00237; GPCRHHODOPSN.  
 CC PROSITE; PS00237; G\_PROTEIN\_RECEP\_F1\_1; 1.  
 CC PROSITE; PS00262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW Chemotaxis.  
 FT DOMAIN 1 35 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 36 58 1 (POTENTIAL).  
 FT DOMAIN 59 69 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 70 91 2 (POTENTIAL).  
 FT DOMAIN 92 108 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 109 129 3 (POTENTIAL).  
 FT DOMAIN 130 148 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 149 170 4 (POTENTIAL).  
 FT DOMAIN 171 216 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 217 237 5 (POTENTIAL).  
 FT DOMAIN 238 253 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 254 277 6 (POTENTIAL).  
 FT DOMAIN 278 296 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 297 316 7 (POTENTIAL).  
 FT DOMAIN 317 364 CYTOPLASMIC (POTENTIAL).





```
DR MIM; 603707; --
DR MIM; 252150; --
DR InterPro; IPR006638; Elp3.
DR InterPro; IPR000385; MaaA NifB PqqE.
DR InterPro; IPR007197; Radical SAM.
DR Pfam; PF04055; Radical SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR PROSITE; PS01305; MAA NIFB PQQE; 1.
KW Molybdenum cofactor biosynthesis; Metal-binding; Iron; Iron-sulfur;
KW 3Fe-4S; Alternative splicing; Disease mutation.
FT METAL 80 80 IRON-SULFUR (3FE-4S) (POTENTIAL).
FT METAL 87 87 IRON-SULFUR (3FE-4S) (POTENTIAL).
FT METAL 312 312 IRON-SULFUR (3FE-4S) (POTENTIAL).
FT METAL 329 329 IRON-SULFUR (3FE-4S) (POTENTIAL).
FT VARSPLIC 1 41 MAAPLSRLRLRLSSARSSGAPVTQPCGESARAASE
FT -> MWKSWKLRDVRREGAGSGSPASSQPSGRGPFPLPGL
FT SSQ (in isoform 2).
FT /FTid=VSP 007395.
FT VARSPLIC 1 41 MAAPLSRLRLRLSSARSSGAPVTQPCGESARAASE
FT -> MWKSWKLRDVR (in isoform 3).
FT /FTid=VSP 007396.
FT VARSPLIC 1 87 Missing (in isoform 4).
FT /FTid=VSP 007397.
FT VARSPLIC 368 385 GMFISQMKNRPMILIGG -> E (in isoform 4).
FT /FTid=VSP 007398.
FT VARIANT 73 73 R -> W (in MOCOD type A).
FT /FTid=VAR 015658.
FT VARIANT 126 126 G -> D (in MOCOD type A).
FT /FTid=VAR 015659.
FT VARIANT 127 127 G -> D (in MOCOD type A).
FT /FTid=VAR 015660.
FT VARIANT 319 319 R -> Q (in MOCOD type A).
FT /FTid=VAR 015661.
FT VARIANT 324 324 G -> E (in MOCOD type A).
FT /FTid=VAR 015662.
FT MUTAGEN 384 385 MISSING: ABOLISHES ACTIVITY.
FT MUTAGEN 384 384 G->S,C: DECREASES ACTIVITY.
FT MUTAGEN 384 384 G->V,D: ABOLISHES ACTIVITY.
FT MUTAGEN 385 385 G->C,P: ABOLISHES ACTIVITY.
FT CONFLICT 239 239 L -> H (IN REF. 1 AND 3).
SQ SEQUENCE 385 AA; 43088 MW; 77AB231D6DBB267E CRC64;

Query Match 48.2%; Score 40; DB 1; Length 385;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 TSAPDTRPAGSTA 16
: |||:|:|:|
Db 23 SGAPVTQPCGESA 36

RESULT 40
ODO2_AZOVI STANDARD; PRT; 398 AA.
AC P20708; Q44474;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dihydrolipoamide succinyltransferase component of 2-oxoglutarate
DE dehydrogenase complex (EC 2.3.1.61) (E2).
GN SUCB OR ODH3.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 478;
RX MEDLINE=90126825; PubMed=2404760;
RA Westphal A.H.; de Kok A.;
RT "The 2-oxoglutarate dehydrogenase complex from Azotobacter
vinelandii. 2. Molecular cloning and sequence analysis of the gene
encoding the succinyltransferase component.";
```

```
RL Eur. J. Biochem. 187:235-239 (1990).
RN [2]
RP SEQUENCE OF 353-398 FROM N.A.
RC STRAIN=ATCC 478;
RX MEDLINE=8816699; PubMed=2832161;
RA Westphal A.H.; de Kok A.;
RT "Lipoamide dehydrogenase from Azotobacter vinelandii. Molecular
cloning, organization and sequence analysis of the gene.";
RL Eur. J. Biochem. 172:299-305 (1988).
RN [3]
RP SEQUENCE OF 1-29 FROM N.A.
RC STRAIN=ATCC 478;
RX MEDLINE=90126823; PubMed=2404759;
RA Schulze B.; Westphal A.H.; Hanemaaijer R.; de Kok A.;
RT "The 2-oxoglutarate dehydrogenase complex from Azotobacter
vinelandii. 1. Molecular cloning and sequence analysis of the gene
encoding the 2-oxoglutarate dehydrogenase component.";
RL Eur. J. Biochem. 187:229-234 (1990).
RN [4]
RP STRUCTURE BY NMR OF 1-79.
RX MEDLINE=96096733; PubMed=8529634;
RA Berg A.; Smits O.; de Kok A.; Vervoort J.;
RT "Sequential 1H and 15N nuclear magnetic resonance assignments and
secondary structure of the lipoyl domain of the 2-oxoglutarate
dehydrogenase complex from Azotobacter vinelandii. Evidence for high
structural similarity with the lipoyl domain of the pyruvate
dehydrogenase complex.";
RL Eur. J. Biochem. 234:148-159 (1995).
RN [5]
RP STRUCTURE BY NMR OF 1-77.
RX MEDLINE=96374493; PubMed=8780784;
RA Berg A.; Vervoort J.; de Kok A.;
RT "Solution structure of the lipoyl domain of the 2-oxoglutarate
dehydrogenase complex from Azotobacter vinelandii.";
RL J. Mol. Biol. 261:432-442 (1996).
CC -!- FUNCTION: THE 2-OXOGLUTARATE DEHYDROGENASE COMPLEX CATALYZES THE
OVERALL CONVERSION OF 2-OXOGLUTARATE TO SUCCINYL-COA & CO(2). IT
CONTAINS MULTIPLE COPIES OF 3 ENZYMAIC COMPONENTS: 2-OXOGLUTARATE
DEHYDROGENASE (E1), DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE (E2) AND
LIPOAMIDE DEHYDROGENASE (E3).
CC -!- CATALYTIC ACTIVITY: Succinyl-CoA + dihydrolipoamide = CoA + S-
succinyl-dihydrolipoamide.
CC -!- COFACTOR: Contains 1 covalently bound lipoyl cofactor.
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBUNIT: Forms a 24-polypeptide structural core with octahedral
symmetry.
CC -!- SIMILARITY: Belongs to the 2-oxoacid dehydrogenase family.
CC -!- SIMILARITY: Contains 1 lipoyl-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M37307; AAA22138.1; ALT_SEQ.
DR EMBL; X52432; CAA36678.1; -.
DR EMBL; X52433; CAA36681.1; -.
DR PIR; S07779; S07779.
DR PDB; 1GHJ; 11-JAN-97.
DR PDB; 1GK; 11-JAN-97.
DR InterPro; IPR001078; 2Oxoacid dh.
DR InterPro; IPR000089; Biotin lipoyl.
DR InterPro; IPR004167; E3 binding.
DR InterPro; IPR003016; Lipoyl_BS.
DR InterPro; IPR006255; SucB.
DR Pfam; PF00198; 2-oxoacid dh; 1.
DR Pfam; PF00364; biotin lipoyl; 1.
DR Pfam; PF02817; e3 binding; 1.
DR ProDom; PD001115; 2Oxoacid dh; 1.
DR TIGRFAMs; TIGR01347; sucB; 1.
```

DR PROSITE; PS00189; LIPOYL; 1.  
KW Tricarboxylic acid cycle; Transferase; Acyltransferase; Lipoyl;  
KW 3D-structure.

FT	INIT MET	0	0
FT	BINDING	42	42
FT	ACT SITE	369	369
FT	ACT SITE	373	373
FT	STRAND	2	5
FT	STRAND	16	18
FT	TURN	25	26
FT	STRAND	28	29
FT	STRAND	34	39
FT	STRAND	44	48
FT	STRAND	53	58
FT	TURN	62	63
FT	STRAND	65	66
FT	TURN	68	69
FT	STRAND	71	75

LIPOYL (POTENTIAL).  
BY SIMILARITY.  
BY SIMILARITY.

SQ SEQUENCE 398 AA; 41871 MW; E5BCA9334123EFB6 CRC64;

Query Match 48.2%; Score 40; DB 1; Length 398;  
Best Local Similarity 61.5%; Pred. No. 94;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 SAPDTRPAPGSTA 16  
Db 145 SAPAGQPAPAATA 157

Search completed: May 6, 2004, 16:29:40  
Job time : 9.61539 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:25:44 ; Search time 29.9487 Seconds  
(without alignments)  
168.565 Million cell updates/sec

Title: US-10-070-566-6

Perfect score: 83

Sequence: 1 GVTSAPDTRPAGSTA 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	60	72.3	20	4	Q9UMI8	Q9umi8 homo sapien
2	59	71.1	553	6	Q9MZL1	Q9mzl1 macaca mula
3	57	68.7	745	16	Q89X06	Q89x06 bradyrhizob
4	57	68.7	1334	16	Q9RKR9	Q9rkr9 streptomyce
5	50	60.2	182	2	Q49921	Q49921 mycobacteri
6	50	60.2	263	16	Q8NMU7	Q8nm7 corynebacte
7	50	60.2	411	16	Q9CCM2	Q9ccm2 mycobacteri
8	49	59.0	175	5	Q23505	Q23505 caenorhabdi
9	49	59.0	260	5	Q23503	Q23503 caenorhabdi
10	49	59.0	1101	16	O53347	O53347 mycobacteri
11	49	59.0	1101	16	Q7TX12	Q7tx12 mycobacteri
12	49	59.0	1765	10	Q7XEI5	Q7xel5 oryza sativ
13	47	56.6	508	11	Q8CD55	Q8cd55 mus musculu
14	47	56.6	533	11	Q7TQE2	Q7tqe2 mus musculu
15	47	56.6	534	2	Q9RBJ1	Q9rbj1 acetobacter
16	47	56.6	633	16	Q9RRT1	Q9rrt1 deinococcus

17	46.5	56.0	367	16	Q88HQ2	Q88hq2 pseudomonas
18	46	55.4	108	17	Q9YEB6	Q9yeb6 aeropyrum p
19	46	55.4	162	6	Q8MJW2	Q8mjw2 equus hemio
20	46	55.4	162	6	Q8MJW4	Q8mjw4 equus asinu
21	46	55.4	168	6	Q8MJV9	Q8mjv9 equus grevy
22	46	55.4	168	6	Q8MJV8	Q8mjv8 equus zebra
23	46	55.4	399	5	Q17326	Q17326 caenorhabdi
24	46	55.4	399	5	Q17358	Q17358 caenorhabdi
25	46	55.4	399	5	Q17588	Q17588 caenorhabdi
26	46	55.4	586	16	Q7UQ15	Q7uql5 rhodopirell
27	46	55.4	889	16	Q9F2N5	Q9f2n5 streptomyce
28	46	55.4	995	5	Q9U7N8	Q9u7n8 manduca sex
29	46	55.4	1751	10	Q7XSY7	Q7xsy7 oryza sativ
30	45	54.2	115	17	Q9YAU7	Q9yau7 aeropyrum p
31	45	54.2	475	16	Q8FQJ1	Q8fqj1 corynebacte
32	45	54.2	879	11	Q7TPK0	Q7tpk0 rattus norv
33	44.5	53.6	635	16	Q9RW96	Q9rw96 deinococcus
34	44	53.0	237	10	Q8LJD3	Q8ljd3 oryza sativ
35	44	53.0	249	11	Q9CV69	Q9cv69 mus musculu
36	44	53.0	273	4	Q96D96	Q96d96 homo sapien
37	44	53.0	294	16	Q8D5A4	Q8d5a4 vibrio vuln
38	44	53.0	341	16	Q89RH8	Q89rh8 bradyrhizob
39	44	53.0	405	16	O53380	O53380 mycobacteri
40	44	53.0	405	16	Q7TWP8	Q7twp8 mycobacteri
41	44	53.0	421	5	Q27393	Q27393 drosophila
42	44	53.0	430	2	Q9LX1	Q9lxl1 acetobacter
43	44	53.0	457	2	Q83W43	Q83w43 pseudomonas
44	44	53.0	484	11	Q9QXX8	Q9qxx8 mus musculu
45	44	53.0	522	16	Q8PKK8	Q8pkk8 xanthomonas

#### ALIGNMENTS

##### RESULT 1

Q9UMI8  
ID Q9UMI8 PRELIMINARY; PRT; 20 AA.  
AC Q9UMI8;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Mucin (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89235154; PubMed=2715633;  
RA King P.X., Tjandra J.J., Reynolds K., McLaughlin P.J., Purcell D.F.J.,  
RA McKenzie I.F.C.;  
RT \*Reactivity of anti-human milk fat globule antibodies with synthetic  
RT peptides.\*;  
RL J. Immunol. 142:3503-3509(1989).  
DR EMBL; M26316; AAA36336.1; -.  
DR PIR; S10571; S10571.  
FT NON TER 1  
FT NON TER 20 20  
SQ SEQUENCE 20 AA; 1887 MW; 5B3473EAEBAFAD87 CRC64;

Query Match 72.3%; Score 60; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.021;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PDTRPAPGSTA 16  
| | | | | | | | | |  
Db 1 PDTRPAPGSTA 11

##### RESULT 2

Q9MZL1  
ID Q9MZL1 PRELIMINARY; PRT; 553 AA.  
AC Q9MZL1;





```
DR ProDom; PD000329; Trans_reg_C; 1.
KW Complete proteome.
SQ SEQUENCE 1334 AA; 138787 MW; 78DC746883E8778C CRC64;
```

Query Match 68.7%; Score 57; DB 16; Length 1334;  
Best Local Similarity 73.3%; Pred. No. 4.9;  
Matches 11: Conservative 0; Mismatches 4; Indels

Qy 1 GVT SAPDTRPAGST 15  
Db 415 GTTPAGGTAPAGST 429

## RESULT 5

Q49921 PRELIMINARY; PRT; 182 AA.  
AC Q49921;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE PbpC.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=i769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Smith D.R.†  
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Robison K.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; UG0022; AAA17345.1; -.  
DR PIR; S73046; S73046.  
DR GO; GO:0004185; P:serine carboxypeptidase activity; IEA.  
DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR001967; Peptidase S11.  
DR Pfam; PF00768; Peptidase S11; I.  
DR SEQUENCE 182 AA; 18094 MW; A528BB3F67CB13F8 CRC64;

Query Match 60.2%; Score 50; DB 2; Length 182;  
Best Local Similarity 61.5%; Pred. No. 7;  
Matches 8: Conservative 2; Mismatches 3: Indels

Qy 1 GVTSA PDTRPAG 13  
| : : | | | | |  
Db 88 GIVTAPDTPPVPG 100

## RESULT 6

Q8NMU7 PRELIMINARY; PRT; 263 AA.  
Q8NMU7;  
DT 01-OCT-2002 (TReMBLrel. 22, Created)  
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE NAD-dependent aldehyde dehydrogenases.  
GN CG52467.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
RA Nakagawa S.;  
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP005281; BAB99860.1; --  
DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
DR GO: GO:0008152; P:metabolism; IEA.  
DR InterPro: IPR002086; Aldehyde dehydr.

DR pfam; PF00171; aldedh; 1.  
KW Complete protecme.  
SQ SEQUENCE 263 AA; 28255 MW; 795CF887B8EC7662 CRC64;

Query Match	60.2%	Score 50;	DB 16;	Length 263;
Beat Local Similarity	62.5%;	Pred. No. 10;		
Matches 10:	Conservative	3;	Mismatches	3;
			Indels	0;
			Gaps	0;

```
QY      1  GVTSAFDTRPAPGSTA 16
        ||||:||||:|:|
Db     101 GVTSSEETOIGEMATA 116
```

## RESULT 7

Q9CCM2  
ID Q9CCM2 PRELIMINARY; PRT; 411 AA.  
AC Q9CCM2;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Putative D-alanyl-D-alanine carboxypeptidase.  
GN ML0691.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Eonore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RL Nature 409:1007-1011(2001).  
DR EMBL; AL583919; CAC30200.1; -.  
DR PIR; D86995; D86995.  
DR Leproma; ML0691; -.  
DR GO; GO:0004185; F:serine carboxypeptidase activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR001967; Peptidase\_S11.  
DR Pfam; PF00768; Peptidase\_S11; I.  
DR PRINTS; PR00725; DADACBPTASE1.  
KW Carboxypeptidase; Complete proteome.  
SQ SEQUENCE 411 AA: 42451 MW: C558FEBB61CF9045 CRC64;

Query Match	60.2%;	Score 50;	DB 16;	Length 411;
Best Local Similarity	61.5%;	Pred. No. 16;		
Matches 8: Conservative	2: Mismatches	3: Indels	0: Gaps	0;

Qy 1 GUTSAPDTRPAPG 13  
| : : | | | | |  
Db 88 GUTAPDTPPVPG 10

## RESULT 8

Q23505	PRELIMINARY;	PRT;	175 AA.
AC	Q23505; Q23506;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Hypothetical protein ZK470.2 in chromosome X.		
GN	ZK470.2.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
QC	Rhabditidae; Pelodierinae; Caenorhabditis.		

```
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RA Minx M.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named isoforms=2;
CC Name=a;
CC IsoId=Q23505-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=b;
CC IsoId=Q23505-2; Sequence=VSP 050209;
CC Note=No experimental confirmation available;
CC EMBL; U39651; AAA80395.1; -.
DR EMBL; U39651; AAA80396.1; -.
DR PIR; T27875; T27875.
DR WormPep; ZK470.2a; CE05086.
DR WormPep; ZK470.2b; CE05087.
KW Hypothetical protein; Alternative splicing.
FT VARSELIC 1 97 Missing (in isoform b).
FT /FTId=VSP 050209.
SQ SEQUENCE 175 AA; 19020 MW; 22E47978F7060ADC CRC64;

Query Match 59.0%; Score 49; DB 5; Length 175;
Best Local Similarity 83.3%; Pred. No. 9.6;
Matches 10; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 3 TSAPDTRPAPGS 14
Db 129 TSAPVTTAPGS 140

RESULT 9
Q23503
ID Q23503 PRELIMINARY; PRT; 260 AA.
AC Q23503;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein ZK470.2c.
GN ZK470.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Minx M.;
RT "The sequence of C. elegans cosmid ZK470.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39651; AAA80393.2; -.
DR PIR; T27873; T27873.
DR WormPep; ZK470.2c; CE31269.
KW Hypothetical protein.
SQ SEQUENCE 260 AA; 29034 MW; 727BCBFE148D3A96 CRC64;

Query Match 59.0%; Score 49; DB 5; Length 260;
Best Local Similarity 83.3%; Pred. No. 14;
```

```
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGS 14
Db 214 TSAPVTTAPGS 225

RESULT 10
O53347
ID O53347 PRELIMINARY; PRT; 1101 AA.
AC O53347;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ATP-dependent DNA helicase (Helicase, UvrD/Rep family).
GN RV3201C OR MT014.45C OR MT3295.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021646; CAAL6666.1; -.
DR EMBL; AE007142; AAK47638.1; -.
DR PIR; G70951; G70951.
DR HSSP; P56255; 1PJR.
DR TIGR; MT3295; -.
DR TuberculList; Rv3201c; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:000628.; P:DNA repair; IEA.
DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 1101 AA; 116687 MW; 269C6AAD3657412B CRC64;

Query Match 59.0%; Score 49; DB 16; Length 1101;
Best Local Similarity 56.2%; Pred. No. 66;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAPGSTA 16
Db 1073 GITVVEDELAPGELA 1088

RESULT 11
Q7TX12
```



RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC054775; A4H54775.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 533 AA; 57027 MW; 1396D0A256AC12CD CRC64;  
  
Query Match 56.6%; Score 47; DB 11; Length 533;  
Best Local Similarity 72.7%; Pred. No. 62;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 6 PDTRPAPGSTA 16  
| : : : : :  
DB 155 PSTKPAPGGTA 165  
  
RESULT 15  
Q9RBJ1 ID Q9RBJ1 PRELIMINARY; PRT; 534 AA.  
AC Q9RBJ1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Levanase precursor.  
GN L5DB.  
CS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
OC Acetobacteraceae; Gluconacetobacter.  
OX NCBI\_TaxID=33996;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SRT4;  
RX MEDLINE=96253999; PubMed=8704949;  
RA Arrieta J., Hernandez L., Coego A., Suarez V., Balmori E.,  
RA Menendez C., Petit-Glatron M.F., Chamber R., Selman-Housein G.;  
RT "Molecular characterization of the levansucrase gene from the  
RT endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4.";  
RL Microbiology 142:1077-1085(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SRT4;  
RX MEDLINE=99373256; PubMed=10441728;  
RA Hernandez L., Arrieta J., Betancourt L., Falcon V., Madrazo J.,  
RA Coego A., Menendez C.;  
RT "Levansucrase from Acetobacter diazotrophicus SRT4 is secreted via  
RT periplasm by a signal-peptide-dependent pathway.";  
RL Curr. Microbiol. 39:146-152(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SRT4;  
RA Menendez C., Hernandez L., Mendoza M.F., Hevia P., Selman G.,  
RA Arrieta J.;  
RT "Molecular cloning and expression in E. coli of an exolevanase from  
RT Acetobacter diazotrophicus SRT4.";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
DR EMBL; L41732; AAF16405.1; -.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR001362; Glyco\_hydro\_32.  
DR Pfam; PF00251; Glyco\_hydro\_32; 1.  
DR SMART; SM00640; Glyco\_32; 1.  
DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.

KW Glycosidase; Hydrolase; Signal.  
FT SIGNAL 1 36 POTENTIAL.  
FT CHAIN 37 534 LEVANASE.  
SQ SEQUENCE 534 AA; 58367 MW; 7BBD5A4A9BA478F6 CRC64;  
  
Query Match 56.6%; Score 47; DB 2; Length 534;  
Best Local Similarity 64.3%; Pred. No. 62;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 GVTSAPDTRPAPGS 14  
| | | | |  
DB 286 GVTFPDPALPPPGS 299  
  
RESULT 16  
Q9RRRT1 ID Q9RRRT1 PRELIMINARY; PRT; 633 AA.  
AC Q9RRRT1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE ABC transporter, ATP-binding protein, MSBA family.  
GN DR2404.  
OS Deinococcus radiodurans.  
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;  
RX MEDLINE=20036396; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey B.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AB002071; AAF11948.1; -.  
DR PIR; E75277; E75277.  
DR TIGR; DR2404; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . . ; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR001140; ABC\_TM\_transpt.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00664; ABC\_membrane; 1.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW ATP-binding; Complete proteome.  
SQ SEQUENCE 633 AA; 69337 MW; DB8549554C15EB32 CRC64;  
  
Query Match 56.6%; Score 47; DB 16; Length 633;  
Best Local Similarity 75.0%; Pred. No. 74;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 GVTSAPDTRPAP 12  
| | | | |  
DB 366 GVTLAPDRPLP 377  
  
RESULT 17  
Q88HQ2 ID Q88HQ2 PRELIMINARY; PRT; 367 AA.  
AC Q88HQ2;

```
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE RND efflux membrane fusion protein.
GN PP3301.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzappple E., Scanlan D., Tran K., Moazzez A.,
RA Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Helm S.,
RA Kiewitz C., Eiser J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AB016786; AAN68908.1; -.
DR TIGR; PP3301; -.
DR InterPro; IPR000437; Prok lipoprot S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 367 AA; 38214 MW; 42EF7CCE920A9B29 CRC64;

Query Match 56.0%; Score 46.5; DB 16; Length 367;
Best Local Similarity 68.8%; Pred. No. 50;
Matches 11; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 2 VTSAPDT-RPAPGSTA 16
D5 | | | | | | | | | |
208 VVYLPETLRPAPGSTA 223

RESULT 18
Q9YEB6 PRELIMINARY; PRT; 108 AA.
AC Q9YEB6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE0658.
GN APE0658.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79630.1; -.
DR PIR; F72653; F72653.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 108 AA; 11371 MW; B4CF6E7DE235A18F CRC64;

Query Match 55.4%; Score 46; DB 17; Length 108;
Best Local Similarity 69.2%; Pred. No. 16;
Matches 11; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
```

```
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVTSAPTTRPAPG 13
D5 | | | | | | | | | |
79 GVGSHPGSRPAPG 91

RESULT 19
Q8MJW2 PRELIMINARY; PRT; 162 AA.
AC Q8MJW2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN DRD4.
OS Equus hemionus (Kulan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=a;
RA Hasegawa T., Ishida N., Sato F., Kusunose R., Oki H.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
RT receptor genes (DRD4) of genus Equus.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB080631; BAC10656.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1
FT NON_TER 162
SQ SEQUENCE 162 AA; 16608 MW; 0C742D20FC4A5C37 CRC64;

Query Match 55.4%; Score 46; DB 6; Length 162;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 VTSAPDTTRPAPGSTA 16
D5 | | | | | | | | | |
87 VTPADATPPPDATA 101

RESULT 20
Q8MJW4 PRELIMINARY; PRT; 162 AA.
AC Q8MJW4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN DRD4.
OS Equus asinus (Donkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9793;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=a;
RA Hasegawa T., Ishida N., Sato F., Kusunose R., Oki H.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
RT receptor genes (DRD4) of genus Equus.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB080629; BAC10654.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1
FT NON_TER 162
SQ SEQUENCE 162 AA; 16677 MW; 56183D31EA5B3A77 CRC64;

Query Match 55.4%; Score 46; DB 6; Length 162;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```



```
QY      2 VTSAPDTRPAPGSTA 16
Db      87 VTPAPDATPPDATA 101

RESULT 21
Q8MJV9
ID Q8MJV9 PRELIMINARY; PRT; 168 AA.
AC Q8MJV9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN DRD4.
OS Equus grevyi (Grevy's zebra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9792;
RN [1]
RP SEQUENCE FROM N.A.
RA Hasegawa T., Ishida N., Sato F., Kusunose R., Oki H.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
RT receptor genes (DRD4) of genus Equus.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB080634; BAC10659.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 168 168
SQ SEQUENCE 168 AA; 17258 MW; FB12046546D0B185 CRC64;

Query Match 55.4%; Score 46; DB 6; Length 168;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 5; Indels 5; Gaps 0;

QY      2 VTSAPDTRPAPGSTA 16
Db      87 VTPAPDATPPDATA 101

RESULT 22
Q8MJV8
ID Q8MJV8 PRELIMINARY; PRT; 168 AA.
AC Q8MJV8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE D4 dopamine receptor (Fragment).
GN DRD4.
OS Equus zebra (Mountain zebra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9791;
RN [1]
RP SEQUENCE FROM N.A.
RA Hasegawa T., Ishida N., Sato F., Kusunose R., Oki H.;
RT "Determination and variability of nucleotide sequences for D4 dopamine
RT receptor genes (DRD4) of genus Equus.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB080635; BAC10660.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 168 168
SQ SEQUENCE 168 AA; 17316 MW; FA0D06B546D0B355 CRC64;

Query Match 55.4%; Score 46; DB 6; Length 168;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 5; Indels 5; Gaps 0;

QY      2 VTSAPDTRPAPGSTA 16
Db      87 VTPAPDATPPDATA 101
```

```
Db      87 VTPAPDATPPDATA 101

RESULT 23
Q17326
ID Q17326 PRELIMINARY; PRT; 399 AA.
AC Q17326;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE N2 basic-helix-loop-helix transcription factor E12/47 homolog.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RA Krause M.W., Fire A.;
RT "cDNA sequence of the C. elegans homolog of the vertebrate basic-
RT helix-loop-helix transcription factor, E12/47.";
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
DR EMBL; U13614; AAA21347.1; -.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50888; HLH_2; 1.
DR PROSITE; PS50888; HLH_2; 1.
SQ SEQUENCE 399 AA; 43225 MW; 10BE9F7FE80F0B80 CRC64;

Query Match 55.4%; Score 46; DB 5; Length 399;
Best Local Similarity 69.2%; Pred. No. 65;
Matches 9; Conservative 1; Mismatches 3; Indels 3; Gaps 0;

QY      3 TSAPDTRPAPGST 15
Db      103 TTPAPDATPPDATA 115

RESULT 24
Q17358
ID Q17358 PRELIMINARY; PRT; 399 AA.
AC Q17358;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transcription factor E12/47 homolog.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RA Krause M.W.;
RT "Genomic sequence of the C. elegans homolog of the vertebrate basic-
RT helix-loop-helix transcription factor, E12/47.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
DR EMBL; U30248; AAC13874.1; -.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50888; HLH_2; 1.
DR PROSITE; PS50888; HLH_2; 1.
SQ SEQUENCE 399 AA; 43210 MW; 2D8F904E90A8CC48 CRC64;

Query Match 55.4%; Score 46; DB 5; Length 399;
Best Local Similarity 69.2%; Pred. No. 65;
Matches 9; Conservative 1; Mismatches 3; Indels 3; Gaps 0;
```

```
QY 3 TSAPDTRPAPGST 15
Db 103 TTAPSTAPAPTST 115

RESULT 25
Q17588
ID Q17588 PRELIMINARY; PRT; 399 AA.
AC Q17588; Q21519;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE M05B5.5 protein.
GN M05B5.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gardner A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Berks M.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
DR EMBL; Z71265; CAA95837.1; -.
DR EMBL; Z71258; CAA95837.1; JOINED.
DR EMBL; Z71258; CAA95787.1; -.
DR EMBL; Z71265; CAA95787.1; JOINED.
DR PIR; T18853; T18853.
DR WormPep; M05B5.5a; CE06191.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00310; HLH; 1.
DR SMART; SMO0353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50888; HLH_2; 1.
SQ SEQUENCE 399 AA; 43192 MW; 8E139D03E597863B CRC64;

Query Match 55.4%; Score 46; DB 5; Length 399;
Best Local Similarity 69.2%; Pred. No. 65;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGST 15
Db 103 TTAPSTAPAPTST 115

RESULT 26
Q7UQ15
ID Q7UQ15 PRELIMINARY; PRT; 586 AA.
AC Q7UQ15;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
```

```
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB6581.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT *Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294144; CAD74891.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 586 AA; 61751 MW; DD2A36C42D4B4EF9 CRC64;

Query Match 55.4%; Score 46; DB 16; Length 586;
Best Local Similarity 61.5%; Pred. No. 97;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAPG 13
Db 349 GMPSPPTAPAPG 361

RESULT 27
Q9F2N5
ID Q9F2N5 PRELIMINARY; PRT; 889 AA.
AC Q9F2N5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein SC03115.
GN SC03115 OR SCE41.24C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3{2};
RA Saunders D.C., Harris D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3 (2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3 (2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3 (2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
```

```

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939115; CAC09556.1; -.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0035201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPOELASTIN.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 889 AA; 88030 MW; 3A85F0B81E3E4F2F CRC64;

Query Match 55.4%; Score 46; DB 16; Length 889;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 SAPDTRPAPGST 15
Db 398 SAPATPPAPGAT 409
|||||:|||||
|||||:|||||

RESULT 28
Q9U7N8 PRELIMINARY; PRT; 995 AA.
AC Q9U7N8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aminopeptidase N (EC 3.4.11.2).
GN APN120.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sphingioidea;
OC Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Midgut;
RA Luo X., McLachlin J.R., Brown M.R., Adarg M.J.;
RT "Expression of a glycosylphosphatidylinositol-linked Manduca sexta
RT aminopeptidase N in insect cells.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF123313; AAF07223.1; -.
DR GO; GO:0004177; F:aminopeptidase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004179; F:membrane alanyl aminopeptidase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001930; Peptidase M1.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Aminopeptidase; Hydrolase.
SQ SEQUENCE 995 AA; 111899 MW; 7B8090CF97D06D11 CRC64;

Query Match 55.4%; Score 46; DB 5; Length 995;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGS 14
Db 963 TEAPATTPAPGS 974
|||||:|||||
|||||:|||||

RESULT 29
Q7XSY7 PRELIMINARY; PRT; 1751 AA.
ID Q7XSY7
```

```

AC Q7XSY7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNB0056F09.9 protein.
GN OSJNB0056F09.9.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,
RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,
RA Chen L., Fan D.L., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,
RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,
RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
RA Zhang R.Q., Guan J.P., Hong G.F.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL606602; CAE01746.1; -.
SQ SEQUENCE 1751 AA; 191607 MW; 034740CF96BB6DC2 CRC64;

Query Match 55.4%; Score 46; DB 10; Length 1751;
Best Local Similarity 60.0%; Pred. No. 3e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VTSAPDTRPAPGSTA 16
Db 1177 VPSAPDPQDGPAGATA 1191
|||||:|||||
|||||:|||||

RESULT 30
Q9YAU7 PRELIMINARY; PRT; 115 AA.
ID Q9YAU7;
AC Q9YAU7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE1847.
GN APE1847.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80851.1; -.
DR PIR; F72570; F72570.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 115 AA; 12474 MW; 9CD70DCCEF89BCC0 CRC64;

Query Match 54.2%; Score 45; DB 17; Length 115;
Best Local Similarity 69.2%; Pred. No. 25;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGST 15
Db 83 TSTFTTRPSPGST 95
|||||:|||||
|||||:|||||
```

```
RESULT 31
Q8FQJ1
ID Q8FQJ1 PRELIMINARY; PRT; 475 AA.
AC Q8FQJ1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN CE1128.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarayayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005217; BAC17938.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 475 AA; 49224 MW; 123C17601AA37490 CRC64;

Query Match 54.2%; Score 45; DB 16; Length 475;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 VTSAPDTRPAPGST 15
Db 372 VTPAPVTTTAPGST 385

RESULT 32
Q7TPK0
ID Q7TPK0 PRELIMINARY; PRT; 879 AA.
AC Q7TPK0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ac2-125.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Xu C.S., Li W.Q., Li Y.C., Chai L.Q., Yuan J.Y., Yang K.J., Yan H.M.,
RA Chang C.F., Zhao L.F., Ma H., Wang L., Wang S.F., Han H.P., Wang G.P.,
RA Shi J.B., Rahman S., Wang Q.N., Zhang J.B.;
RT "Liver regeneration after PH."
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY321335; AAP86267.1; -.
SQ SEQUENCE 879 AA; 97518 MW; F95A2E0A3645C590 CRC64;

Query Match 54.2%; Score 45; DB 11; Length 879;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGS 14
Db 683 SSTPSTRPPGS 694

RESULT 33
Q9RW96
ID Q9RW96 PRELIMINARY; PRT; 635 AA.
AC Q9RW96;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein DR0773.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL; AB001932; AAF10356.1; -.
DR PIR; F75477; F75477.
DR TIGR; DR0773; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 635 AA; 61390 MW; 57E4A112B1E18640 CRC64;

Query Match 53.6%; Score 44.5; DB 16; Length 635;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 10; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 3 TSAPDTRPAPGSTA 16
Db 421 TAAP-TRPSPSTA 433

RESULT 34
Q8LJD3
ID Q8LJD3 PRELIMINARY; PRT; 237 AA.
AC Q8LJD3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE P0496H05.16 protein.
GN P0496H05.16.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone:P0496H05."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003267; BAC03288.1; -.
DR Gramene; Q8LJD3; -.
SQ SEQUENCE 237 AA; 25649 MW; 704186DD4C9C4E8A CRC64;

Query Match 53.0%; Score 44; DB 10; Length 237;
Best Local Similarity 64.3%; Pred. No. 75;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 TSAPDTRPAPGSTA 16
Db 163 SSPPSRPAPPSTA 176

RESULT 35
Q9CV69
ID Q9CV69 PRELIMINARY; PRT; 249 AA.
```

```
AC Q9CV69;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nuclear FMRP interacting protein (Fragment).
GN NUFIP1 OR NUFIP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DR MGD; MGI:1351474; Nufip1.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003723; F:RNA binding; IDA.
DR InterPro; IPR007087; Znf C2H2.
DR Pfam; PF00096; zf-C2H2; 1.
DR SMART; SM00355; Znf C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
KW Metal-binding; Zinc; Zinc-finger.
FT NON TER 249
SQ SEQUENCE 249 AA; 28272 MW; 2477016BE7415413 CRC64;

Query Match 53.0%; Score 44; DB 11; Length 249;
Best Local Similarity 72.7%; Pred. No. 79;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 SAPDTRPAPGS 14
Db 11 SAPDLTPAPGT 21
|||||

RESULT 36
Q96D96
ID Q96D96 PRELIMINARY; PRT; 273 AA.
AC Q96D96;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to hypothetical protein MGC15619.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; BC009731; AAH09731.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005261; F:cation channel activity; IEA.
DR GO; GO:0006812; P:cation transport; IEA.
DR InterPro; IPR005820; M+channel_nlg.
KW Hypothetical protein.
SQ SEQUENCE 273 AA; 31683 MW; 0F93B428AECBEC4F CRC64;

Query Match 53.0%; Score 44; DB 4; Length 273;
Best Local Similarity 61.5%; Pred. No. 87;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVTSPDTRPAPG 13
Db 67 GRAAPDVAPAPG 79
|||||

RESULT 37
Q8D5A4
ID Q8D5A4 PRELIMINARY; PRT; 294 AA.
AC Q8D5A4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Phosphotransferase system,
DE mannose/fructose/N-acetylglactosamine-specific component IID.
GN WV21020.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016811; AAC07928.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.
DR InterPro; IPR004704; IID_comp.
DR Pfam; PF03613; IID-AGA; 1.
DR Transferase; Complete proteome.
KW SEQUENCE 294 AA; 31874 MW; 2258450D3CF2391B CRC64;

Query Match 53.0%; Score 44; DB 16; Length 294;
Best Local Similarity 72.7%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 PDTRPAPGSTA 16
Db 16 PDVRPAPGVAA 26
|||||

RESULT 38
Q89RH8
ID Q89RH8 PRELIMINARY; PRT; 341 AA.
AC Q89RH8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transcriptional regulatory protein.
GN BLL2794.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:21:39 ; Search time 59.5641 Seconds  
(without alignments)  
109.103 Million cell updates/sec

Title: US-10-070-566-7

Perfect score: 127

Sequence: 1 PDTRPAGSTAPPAHGVTSPDPT 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 29Jan04: \*  
1: Geneseqp1980s: \*  
2: Geneseqp1990s: \*  
3: Geneseqp2000s: \*  
4: Geneseqp2001s: \*  
5: Geneseqp2002s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	100.0	23	4 AAU00489	Aau00489 Human muc
2	127	100.0	24	2 AAY32507	Aay32507 Synthetic
3	127	100.0	28	2 AAR05170	Aar05170 Epitopic
4	127	100.0	28	2 AAW03362	Aaw03362 Mucin tan
5	127	100.0	28	2 AAR95510	Aar95510 Polymorph
6	127	100.0	30	2 AAW35738	Aaw35738 Mucin pep
7	127	100.0	30	2 AAW35737	Aaw35737 Mucin pep
8	127	100.0	30	4 AAB72472	Aab72472 Fusion pr
9	127	100.0	30	4 AAE09543	Aae09543 Human muc
10	127	100.0	31	2 AAW31695	Aaw31695 Mucin pep
11	127	100.0	31	2 AAW31696	Aaw31696 Mucin pep
12	127	100.0	31	6 ABG73816	Abg73816 Tumour-as
13	127	100.0	32	4 AAE09541	Aae09541 Human muc
14	127	100.0	40	2 AAR68002	Aar68002 Mucin rep
15	127	100.0	40	2 AAW54873	Aaw54873 Carcinoma
16	127	100.0	40	2 AAW72703	Aaw72703 Human muc
17	127	100.0	40	5 ABP56039	Abp56039 Repeat mo
18	127	100.0	41	3 AAY96170	Aay96170 MUC1 repe
19	127	100.0	43	6 AAE333952	Aae333952 Dilipidat
20	127	100.0	43	7 ADD88870	Add88870 Synthetic
21	127	100.0	46	5 ABP56035	Abp56035 Histidine
22	127	100.0	46	6 AAE333936	Aae333936 MUC 1 pep
23	127	100.0	50	2 AAW35739	Aaw35739 Mucin pep
24	127	100.0	51	2 AAW31697	Aaw31697 Mucin pep
25	127	100.0	100	5 ABB76181	Abb76181 Synthetic

26	127	100.0	105	2 AAR68022	Aar68022 Mucin pep
27	127	100.0	105	2 AAW72697	Aaw72697 Synthetic
28	127	100.0	216	3 AAY92665	Aay92665 MUC-1 ana
29	127	100.0	508	2 AAW72333	Aaw72333 MiniMUC1
30	127	100.0	1255	5 ABP56040	Abp56040 MUC1 rece
31	127	100.0	1255	6 ABB82568	Abb82568 MUC1/REP
32	127	100.0	1255	6 ABR47537	Abr47537 Breast ca
33	127	100.0	1255	6 ABR92124	Abr92124 Human cer
34	127	100.0	1255	7 ADD45111	Add45111 Human pro
35	127	100.0	1255	7 ADE54622	Ade54622 Human pro
36	122	96.1	173	3 AAY71021	Aay71021 Human muc
37	122	96.1	295	3 AAY71027	Aay71027 Ubiquitin
38	122	96.1	455	3 AAY71024	Aay71024 Human muc
39	122	96.1	473	4 AAB09508	Aae09508 Human muc
40	122	96.1	475	4 AAU00573	Aau00573 Human MUC
41	122	96.1	475	5 ABB77476	Abb77476 Human MUC
42	122	96.1	475	6 ADA50567	Ada50567 Mucin 1 (
43	122	96.1	475	6 AAB37800	Aae37800 Human muc
44	122	96.1	475	7 ADD14120	Add14120 Human src
45	122	96.1	475	7 ADE48133	Ade48133 MUC1 amin

## ALIGNMENTS

### RESULT 1

AAU00489

ID AAU00489 standard; peptide; 23 AA.

XX AC AAU00489;

XX DT 04-JUL-2001 (first entry)

XX XX Human mucin peptide MUC1(23).

XX KW Human; polymorphic epithelial cell mucin; PEM; chimeric virus particle;  
KW CVP; plant virus coat protein; comovirus; CPMV; cowpea mosaic virus;  
KW SBMV; Southern bean mosaic virus; LHSV; red clover necrotic mosaic virus;  
KW RCNMV; MUC1; tumour; cancer vaccine.

XX OS Homo sapiens.

XX PN WO200118199-A1.

XX PD 15-MAR-2001.

XX PF 11-SEP-2000; 2000WO-GB003500.

XX PR 09-SEP-1999; 99GB-00021337.

XX PA (DOWC ) DOW CHEM CO.

XX PI Bendig MM, Jores TD, Longstaff M, Hellendoorn K;

XX XX WPI; 2001-244570/25.

XX PT New chimeric plant virus particles with an immunogenically active peptide  
PT of a tumor-associated mucin, useful as a vaccine or for the manufacture  
PT of a vaccine for treating and/or preventing tumors and/or cancer.

XX PS Claim 7; Page 15; 63pp; English.

XX CC The present sequence representing human mucin peptide MUC1(23) is 1 of 7  
CC polymorphic epithelial cell mucin (PEM) peptides (AAU00483-AAU00489). A  
CC novel chimeric virus particle (CVP) is generated from a plant virus coat  
CC protein with a beta barrel structure and modified by insertion of an  
CC immunogenically active mucin peptide at an immunogenic site in the coat  
CC protein. The mucin peptide epitopes are preferably inserted into the coat  
CC protein of a plant virus such as the comovirus CPMV (cowpea mosaic  
CC virus). The ability of the chimeric virus particle CPMV-MUC1(16) to  
CC elicit antibodies, which can cause regression of tumours expressing the  
CC MUC1 protein, is demonstrated in a mouse tumour model. Other examples of  
CC CVPs include the insertion of the mucin MUC1(16) peptide into the coat

CC proteins of other plant viruses such as SBMV (Southern bean mosaic virus;  
 CC AAU00490), LTVS (AAU00491) and RCNMV (red clover necrotic mosaic virus;  
 CC AAU00492). The CVP is useful as or as part of a vaccine particularly for  
 CC treating and preventing tumours and cancer. The CVP provides advantages  
 CC over prior art antigen-presenting means since conventional live animal  
 CC virus vectors can be avoided; as can the need for separate mucin peptide  
 CC synthesis and chemical-coupling to a conventional carrier. Also, the CVP  
 CC is shown to induce good mucosal immunity, and does not require the  
 CC addition of exogenous adjuvants to induce a strong immune response  
 CC  
 XX Sequence 23 AA;

Query Match 100.0%; Score 127; DB 4; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 9.9e-09;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 Db 1 PDTRPAPGSTAPPAGHGVTSAPDT 23

RESULT 2  
 AAY32507  
 ID AAY32507 standard; peptide; 24 AA.  
 AC AAY32507;

DT 20-OCT-1999 (first entry)  
 DE Synthetic MUC1-VNTR peptide P24.

KW MUC1; VNTR; variable number of tandem repeats; tumour-associated antigen;  
 KW immunodominant peptide; mammary tumour; mucin antigen; T cell activation;  
 KW tumour proliferation inhibitor; carcinoma cell; breast cancer.

OS Synthetic.

PN WO9940881-A2.

PD 19-AUG-1999.

PF 12-FEB-1999; 99WO-EP000941.

PR 13-FEB-1998; 98EP-00102529.

PA (BAST/) BASTERT G.

PI Bastert G, Kaul S;

DR WPI; 1999-508570/42.

PS New antibodies specific for mammary tumor-associated mucin antigen, used  
 to, e.g. inhibit proliferation of tumor.

Example 1; Page 28; 67pp; English.

CC This sequence represents synthetic peptide of mammary tumour associated  
 CC antigen MUC1 variable number of tandem repeats (VNTR) structure. The  
 CC invention relates to an antibody which specifically binds to the  
 CC carbohydrate structure of the mammary tumour-associated mucin antigen  
 CC (MUC1) tandem repeat in carcinoma cells. The antibody: (a) has an  
 CC affinity for a 200-400 kDa glycoprotein fraction from tumour cell-  
 CC containing ascites of breast cancer patients that is at least 100-fold  
 CC that of its affinity for native MUC1 antigen (400-440 kDa) from normal  
 CC cells; (b) does not bind to non-glycosylated MUC1 antigen; and (c)  
 CC binding to the 200-400 kDa glycoprotein changes by 10% if the fraction is  
 CC treated with neuraminidase to cleave N-terminal neuraminic acids, or with  
 CC formalin. The antibody is used to inhibit proliferation of tumour,  
 CC particularly carcinoma cells in a patient, to influence the activity of a  
 CC tumour patient's T cells, to determine tumour cells in a patient's sample  
 CC and to differentiate tumour and normal cells. The antibodies are  
 CC particularly for use with breast cancer patients. The antibodies are more  
 CC specific than those of prior art

XX Sequence 24 AA;  
 SQ  
 Query Match 100.0%; Score 127; DB 2; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 1e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 Db 2 PDTRPAPGSTAPPAGHGVTSAPDT 24

RESULT 3  
 AAR05170  
 ID AAR05170 standard; protein; 28 AA.

AC AAR05170;

DT 25-MAR-2003 (revised)

DT 10-OCT-1999 (first entry)

DE Epitopic fragment of polymorphic epithelial mucin (PEM).

KW Breast cancer; cancer; polymorphic epithelial mucin; mucin; PEM.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Domain 5..24

FT /label= Tandem repeat

PN WO9005142-A.

PD 17-MAY-1990.

PF 10-NOV-1988; 88GB-00026362.

PR 10-NOV-1988; 88GB-00026362.

PR 05-APR-1989; 89GB-00007660.

XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.

PI Taylorpapa J, Gendler S, Burchell J;

DR WPI; 1990-178797/23.

PT Polypeptide(s) having sequence of antigenic epitope on PEM - (polymorphic  
 PT epithelial mucin), useful for producing antibodies for diagnosis and  
 PT therapy of breast cancer.

PS Disclosure; Page ?; -pp; English.

CC The tandem repeat is an epitope in PEM which is expressed at high levels  
 CC in sufferers from breast cancer, and is thus useful in generating  
 CC antibodies for diagnosis and therapy. (Updated on 25-MAR-2003 to correct  
 CC PR field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated on 25-  
 CC MAR-2003 to correct PI field.)

XX Sequence 28 AA;

Query Match 100.0%; Score 127; DB 2; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 Db 5 PDTRPAPGSTAPPAGHGVTSAPDT 27

RESULT 4  
 AAW03362

ID AAW03362 standard; peptide; 28 AA.

XX

AC AAM03362;  
 DT 20-MAR-1997 (first entry)  
 DE Mucin tandem repeat, ratchet library template.  
 DE Cytotoxic T lymphocyte; CTL; epitope; mucin; ratchet library; cancer;  
 KW pharmaceutical; vaccine; treatment; prevention; disease; malignancy;  
 KW tandem repeat; template.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Region 1..20  
 FT /note= "non-MHC restricted CTL epitope"  
 XX  
 PN WO9622067-A2.  
 PN 25-JUL-1996.  
 PD 15-DEC-1995; 95WO-US016290.  
 PF 27-DEC-1994; 94US-00366332.  
 PR (UNBI-) UNITED BIOMEDICAL INC.  
 PA Kuebler PJ, Nixon DF;  
 PI WPI; 1996-354273/35.  
 XX Ratchet library of peptide(s) contg. an immuno-stimulatory CTL epitope -  
 XX derived from longer template peptide, useful as pharmaceutical or vaccine  
 XX against infectious disease or malignancy.  
 XX  
 PS Claim 9; Page 36; 60pp; English.  
 XX  
 CC The present peptide contains a mucin tandem repeat, which elicits non-MHC  
 CC restricted cytotoxic T lymphocyte responses, and can be used as a  
 CC template in the prep. of a ratchet library, comprising peptides contg.  
 CC at least 1 immunostimulatory CTL epitope. Basically the distribution of  
 CC amino acids at each position in the template is calculated, a ratchet  
 CC library constructed from the longer template peptide by sequentially  
 CC ratcheting it into the shorter ratchet length and the peptides  
 CC synthesised using standard solid phase methods. The library can be used  
 CC in pharmaceuticals and vaccines for the treatment, and/or prevention of  
 CC disease and malignancy associated with mucin secretion, e.g. cancer.  
 CC Several epitopes can be incorporated into the same library, rather than  
 CC using a mixt. of individually synthesised immunogenic peptides, which  
 CC helps to overcome problems of genetic diversity and MHC restriction. The  
 CC library may also include antigenic variations and escape mutations  
 XX  
 SQ Sequence 28 AA;  
 Query Match 100.0%; Score 127; DB 2; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 Db 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 RESULT 5  
 AAR95510  
 ID AAR95510 standard; peptide; 28 AA.  
 XX AAR95510;  
 AC  
 XX 06-NOV-1996 (first entry)  
 DT Polymorphic epithelial mucin tumour antigen epitope.  
 DE Abtide; prostate specific mucin antigen; human prostate cancer; LNCaP;  
 XX

KW diagnostic; detection; imaging; tumour; phage; peptide library; breast;  
 KW polymorphic; epithelial.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9609411-A1.  
 XX 28-MAR-1996.  
 PD 20-SEP-1995; 95WO-US011934.  
 PF 21-SEP-1994; 94US-00310192.  
 PR 07-JUN-1995; 95US-00488161.  
 XX (CYTO-) CYTOGEN CORP.  
 PA Alvarez VL;  
 XX WPI; 1996-188471/19.  
 DR New isolated peptide(s) with specific binding activities - obtd. by  
 PT screening random peptide libraries, for use in diagnostic and therapeutic  
 PT compsns.  
 XX Claim 40; Page 90; 106pp; English.  
 PS AAR95510 is an epitope of the polymorphic epithelial tumour antigen found  
 XX on human breast cancer cells. The monoclonal antibody SM-3 is specific  
 CC for this epitope. The epitope is used to produce peptide libraries for  
 CC the isolation of antigen binding peptides (abtides). The abtides are  
 CC identified from random peptide libraries using specific ligand binding.  
 CC Abtides mimic the binding specificity of large molecules such as  
 CC antibodies and receptors but have a much smaller size allowing their  
 CC production at a lower cost and reducing the extent of their  
 CC immunogenicity aiding in vivo delivery. The abtides are useful for the  
 CC diagnosis, detection, imaging and treatment of disease, e.g. tumours,  
 CC prostate cancer and breast cancer  
 XX Sequence 28 AA;  
 Query Match 100.0%; Score 127; DB 2; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 Db 5 PDTRPAPGSTAPPAGHGVTSAPDT 27  
 RESULT 6  
 AAR95510  
 ID AAR95510 standard; peptide; 30 AA.  
 XX AAR95510;  
 AC  
 XX 02-APR-1998 (first entry)  
 DT Mucin peptide MUC1-4.  
 XX MUC1; mucin; vaccine; keyhole limpet haemocyanin; KHL; cancer;  
 KW immunogenic protein; immune response.  
 XX Synthetic.  
 OS Mammalia.  
 XX WO9734921-A1.  
 PN 25-SEP-1997.  
 PD 20-MAR-1997; 97WO-US004493.  
 PF 20-MAR-1996; 96US-0013775P.  
 PR  
 XX





RESULT 9  
AAE09543  
ID AAE09543 standard; peptide; 30 AA.  
XX  
AC AAE09543;  
XX  
DT 19-NOV-2001 (first entry)  
XX  
DE Human mucin-1 (MUC-1) extracellular epitope #14 for CTL.  
XX  
KW Mucin-1; cytostatic; immunostimulant; cell mediated immune response;  
KW carcinoma; adenocarcinoma; breast cancer; dendritic cell; vaccine;  
KW gene therapy; human; MUC-1; cytotoxic T-lymphocyte; CTL.  
XX  
OS Homo sapiens.  
XX  
PN WO200157068-A1.  
XX  
PD 09-AUG-2001.  
XX  
PF 01-FEB-2001; 2001WO-AU000090.  
XX  
PR 01-FEB-2000; 2000AU-00005369.  
XX  
PR 14-JUN-2000; 2000US-00593870.  
XX  
PA (AUST-) AUSTIN RES INST.  
XX  
XX McKenzie IFC, Pietersz GA, Apostolopoulos V;  
PI  
DR WPI; 2001-541537/60.  
XX  
XX Immunostimulant peptide, used as an anti-carcinoma vaccine, comprises a  
PT an epitope of the non-VNTR, non-leader region of a mucin.  
PT  
XX  
PS Disclosure; Page 34; 84pp; English.  
XX  
CC The patent discloses peptide or polypeptides capable of eliciting an  
CC immune response, comprising an amino acid sequence corresponding to an  
CC epitope of the non-central portion of varying numbers of an amino acid  
CC motif (VNTR), non-leader region of a mucin. The peptides of the  
CC invention, fusion proteins comprising the peptide and conjugate compounds  
CC with carbohydrate polymers are used to induce a cell mediated immune  
CC response against mucin in the prevention or treatment of carcinoma,  
CC preferably adenocarcinoma, most preferably breast cancer. They are also  
CC used to pulse dendritic cell for in vivo transfer and use as a vaccine.  
CC They are also used in gene therapy. The present sequence is an  
CC extracellular peptide of mucin-1 (MUC-1) protein from human. This  
CC sequence is an epitope for cytotoxic T-lymphocytes (CTL)  
XX  
SQ Sequence 30 AA;  
  
Query Match 100.0%; Score 127; DB 4; Length 30;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
  
RESULT 10  
AAW31695  
ID AAW31695 standard; peptide; 31 AA.  
XX  
AC AAW31695;  
XX  
DT 02-APR-1998 (first entry)  
XX  
DE Mucin peptide MUC1-3 with C-terminal cysteine.  
XX  
KW MUC1; mucin; vaccine; keyhole limpet haemocyanin; KHL; cancer;  
KW immunogenic protein; immune response; conjugate.  
XX  
PI

OS Synthetic.  
OS Mammalia.  
XX  
PN WO9734921-A1.  
XX  
PD 25-SEP-1997.  
XX  
PF 20-MAR-1997; 97WO-US004493.  
XX  
PR 20-MAR-1996; 96US-0013775P.  
XX  
PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
PI Livingston PO, Zhang S;  
XX  
DR WPI; 1997-480158/44.  
XX  
PT Vaccine effective against cancer of the breast, prostate, colon, lung or  
PT pancreas - comprising mucin peptide, especially MUC1, conjugated to  
PT immunogenic protein, especially keyhole limpet haemocyanin.  
XX  
PS Disclosure; Page 11; 45pp; English.  
XX  
CC This mucin peptide is used in a vaccine capable of producing an immune  
CC response which recognises a mucin. The mucin peptide is selected from  
CC MUC1 peptide group. The vaccine comprises an amount of the mucin peptide  
CC conjugated to an immunogenic protein effective to stimulate or enhance  
CC immune response in the subject, together with an adjuvant and a vehicle.  
CC A cysteine is added to the C-terminal of this peptide to facilitate the  
CC conjugation with protein carriers. The immunogenic protein is a keyhole  
CC limpet haemocyanin (KLH) or its derivative. The vaccine can be used to  
CC induce an immune response in patients suffering from a cancer of the type  
CC where the cancer cells have mucin on their surface, e.g. breast cancer,  
CC prostate cancer, lung cancer, colon cancer or pancreas cancer  
XX  
SQ Sequence 31 AA;  
  
Query Match 100.0%; Score 127; DB 2; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 2 PDTRPAPGSTAPPAGHGVTSAPDT 24  
  
RESULT 11  
AAW31696  
ID AAW31696 standard; peptide; 31 AA.  
XX  
AC AAW31696;  
XX  
DT 02-APR-1998 (first entry)  
XX  
DE Mucin peptide MUC1-4 with N-terminal cysteine.  
XX  
KW MUC1; mucin; vaccine; keyhole limpet haemocyanin; KHL; cancer;  
KW immunogenic protein; immune response; conjugate.  
XX  
OS Synthetic.  
OS Mammalia.  
XX  
PN WO9734921-A1.  
XX  
PD 25-SEP-1997.  
XX  
PF 20-MAR-1997; 97WO-US004493.  
XX  
PR 20-MAR-1996; 96US-0013775P.  
XX  
PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
PI Livingston PO, Zhang S;

XX WPI; 1997-480158/44.

XX Vaccine effective against cancer of the breast, prostate, colon, lung or

PT pancreas - comprising mucin peptide, especially MUC1, conjugated to

PT immunogenic protein, especially keyhole limpet haemocyanin.

XX Disclosure; Page 11; 45pp; English.

XX This mucin peptide is used in a vaccine capable of producing an immune

CC response which recognises a mucin. The mucin peptide is selected from

CC MUC1 peptide group. The vaccine comprises an amount of the mucin peptide

CC conjugated to an immunogenic protein effective to stimulate or enhance

CC immune response in the subject, together with an adjuvant and a vehicle.

CC A cysteine is added to the N-terminal of this peptide to facilitate the

CC conjugation with protein carriers. The immunogenic protein is a keyhole

CC limpet haemocyanin (KLH) or its derivative. The vaccine can be used to

CC induce an immune response in patients suffering from a cancer of the type

CC where the cancer cells have mucin on their surface, e.g. breast cancer,

CC prostate cancer, lung cancer, colon cancer or pancreas cancer

XX Sequence 31 AA;

SQ

Query Match 100.0%; Score 127; DB 2; Length 31;

Best Local Similarity 100.0%; Pred. No. 1.3e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23

Db 6 PDTRPAPGSTAPPAGHGVTSAPDT 28

RESULT 12

ABG73816

ID ABG73816 standard; peptide; 31 AA.

AC ABG73816;

XX 20-MAR-2003 (first entry)

DT

DE Tumour-associated antigen mucin-1 (MUC1) peptide, E31-biotin.

XX Human; MUC1; mucin-1; cytostatic; antagonist; vaccine; tumour;

KW tumour-associated mucin-1; cancer; breast carcinoma; colon carcinoma;

KW oesophageal squamous cell carcinoma; pancreatic carcinoma;

KW prostate carcinoma; multiple myeloma; adenocarcinoma.

XX Homo sapiens.

OS

EH Key Location/Qualifiers

FT 1. .8

FT Region /label= Type\_1\_turn

FT 2. .6

FT Region /label= T\_cell\_epitope

FT 3. .8

FT Region /label= Alt-1 epitope

FT /note= "Alt-1 is a TSA-specific murine monoclonal

FT antibody. Residues 3-8 specifically claimed in claim 34"

FT 7. .12

FT Region /label= Epitope\_region

FT 9. .28

FT Region /label= MUC1\_extracellular\_domain\_core

FT /note= "MUC1 extracellular domain consists of 30 to 90

FT tandem repeats of this sequence"

FT 11. .16

FT Region /label= Epitope\_region

FT 15. .20

FT Region /label= Epitope\_region

FT 19. .24

FT Region /label= Epitope\_region

FT 20. .26

FT Region /label= B\_cell\_epitope

FT 21. .28

FT Region

FT Region /label= Type\_1\_turn

FT 22. .26

FT Region /label= T\_cell\_epitope

FT 23. .28

FT /label= Alt-1 epitope

FT /note= "Alt-1 is a TSA-specific murine monoclonal

FT antibody. Residues 3-8 specifically claimed in claim 34"

FT Modified-site 31

FT /note= "Biotinylated"

XX

PN US2002132771-A1.

XX

PD 19-SEP-2002.

XX

PF 26-NOV-2001; 2001US-00994466.

XX

PR 18-AUG-1999; 99US-0149492P.

PR 11-NOV-1999; 99US-0164714P.

PR 18-AUG-2000; 2000US-00641833.

PR 28-NOV-2000; 2000US-00724094.

PR 21-FEB-2001; 2001US-0270456P.

PR 21-FEB-2001; 2001US-0270471P.

XX

PA (MADI/) MADIYALAKAN R.

XX

PI Madiyalakan R;

XX

DR WPI; 2003-155898/15.

XX

PT New therapeutic compositions comprising a binding agent that binds to

PT tumour-associated MUC1 epitope, useful for treating human tumors, e.g.

PT breast carcinoma, prostate carcinoma or multiple myeloma.

XX

PS Example 24; Page 13; 27pp; English.

XX

CC The invention relates to new therapeutic compositions, which comprise a

CC binding agent that specifically binds to an epitope of a tumour-

CC associated mucin-1 (MUC1), are effective for treating a mammal bearing a

CC tumour. Mice were implanted with 413BCR tumour cells 2 weeks after the

CC start of the immunisation series (using either a conjugate of the binding

CC agent, or a complex of the binding agent-MUC1). It was found that a

CC humoral response was induced in mice treated with both the conjugated and

CC complexed binding agent. A T2 cellular response to the binding agent was

CC induced in these mice. A trend for reduction in tumour mass and size in

CC mice treated with the conjugated or complexed binding agent was also

CC demonstrated. The therapeutic compositions or the method is useful for

CC treating a mammal (particularly a human) bearing a tumour, especially a

CC tumour that expresses a tumour-associated MUC-1. In particular, the

CC compositions are useful for treating adenocarcinomas, e.g. breast

CC carcinoma, colon carcinoma, oesophageal squamous cell carcinoma,

CC pancreatic carcinoma, prostate carcinoma, or multiple myeloma. The

CC present sequence represents the mucin-1 peptide, E31-biotin

XX

SQ Sequence 31 AA;

Query Match 100.0%; Score 127; DB 6; Length 31;

Best Local Similarity 100.0%; Pred. No. 1.3e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRFAPGSTAPPAGHGVTSAPDT 23

Db 2 PDTRFAPGSTAPPAGHGVTSAPDT 24

RESULT 13

AAE09541

ID AAE09541 standard; peptide; 32 AA.

XX

AC AAE09541;

XX

DT 19-NOV-2001 (first entry)

XX

DE Human mucin-1 (MUC-1) VNTR peptide #5.

XX Mucin-1; cytostatic; immunostimulant; cell mediated immune response;  
 KW carcinoma; adenocarcinoma; breast cancer; dendritic cell; vaccine;  
 KW gene therapy; human; MUC-1.  
 OS Homo sapiens.  
 XX WO200157063-A1.  
 EN 09-AUG-2001.  
 PD 01-FEB-2001; 2001WO-AU0000090.  
 XX 01-FEB-2000; 2000AU-00005369.  
 PR 14-JUN-2000; 2000US-00593870.  
 XX (AUST-) AUSTIN RES INST.  
 PA Mckenzie IFC, Pietersz GA, Apostolopoulos V;  
 PI WPI; 2001-541537/60.  
 XX Immunostimulant peptide, used as an anti-carcinoma vaccine, comprises a  
 PT an epitope of the non-VNTR, non-leader region of a mucin.  
 PT Disclosure; Page 34; 84pp; English.  
 PS The patent discloses peptide or polypeptides capable of eliciting an  
 CC immune response, comprising an amino acid sequence corresponding to an  
 CC epitope of the non-central portion of varying numbers of an amino acid  
 CC motif (VNTR), non-leader region of a mucin. The peptides of the  
 CC invention, fusion proteins comprising the peptide and conjugate compounds  
 CC with carbohydrate polymers are used to induce a cell mediated immune  
 CC response against mucin in the prevention or treatment of carcinoma,  
 CC preferably adenocarcinoma, most preferably breast cancer. They are also  
 CC used to pulse dendritic cell for in vivo transfer and use as a vaccine.  
 CC They are also used in gene therapy. The present sequence is a VNTR  
 CC peptide of mucin-1 (MUC-1) protein from human  
 XX Sequence 32 AA;  
 SQ  
 Query Match 100.0%; Score 127; DB 4; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 Db 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 RESULT 14  
 AAR68002  
 ID AAR68002 standard; peptide; 40 AA.  
 AC AAR68002;  
 XX 25-MAR-2003 (revised)  
 DT 05-SEP-1995 (first entry)  
 XX Mucin repeat sequence.  
 DE Mucin; multiple tandem repeat; vaccine; pancreas cancer; breast cancer;  
 KW ovary cancer; colon cancer; HIV.  
 XX Synthetic.  
 OS Key Location/Qualifiers  
 XX Region 3..5  
 FT /label= DTR motif  
 FT /note= "can be substituted by viral sequence, tumor  
 FT antigen or autoantigen"  
 XX WO9503825-A1.  
 PN

XX 09-FEB-1995.  
 PD 29-JUL-1994; 94WO-US008477.  
 PF 30-JUL-1993; 93US-00099354.  
 XX (FINN/) FINN O J.  
 PA (FONT/) FONTENOT J D.  
 PA (MONT/) MONTELA RO R C.  
 XX Finn OJ, Fontenot JD, Montelaro RC;  
 PI WPI; 1995-082033/11.  
 XX Synthetic multiple tandem repeat mucin-1 peptides and analogues - have  
 PT native conformation in the absence of glycosylation and are linked to  
 PT epitopes; for vaccines and tests of cancer, viruses and bacteria.  
 XX Disclosure; Page 32; 125pp; English.  
 PS A synthetic peptide comprises at least 4 tandem repeats of the mucin  
 XX peptide given in (AAR68002). The multiple prolines of the repeat  
 CC structure are needed to maintain a rigid structure. The DTR motif,  
 CC located between the first 2 prolines in each repeat, is the target of an  
 CC anti-mucin immune response, and can be substituted by a sequence from a  
 CC virus, tumor antigen or autoantigen, for use in vaccine development.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX Sequence 40 AA;  
 SQ  
 Query Match 100.0%; Score 127; DB 2; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 Db 2 PDTRPAPGSTAPPAGHGVTSAPDT 24  
 RESULT 15  
 AAW54873  
 ID AAW54873 standard; peptide; 40 AA.  
 XX AAW54873;  
 AC 24-SEP-1998 (first entry)  
 DT Carcinoma-associated antigen DF3/MUC1 tandem repeat segment.  
 XX Mucin; DF3/MUC1; cytotoxic T lymphocyte; autoimmune disease;  
 KW granuloma formation; transplant rejection.  
 XX Homo sapiens.  
 OS WO9817300-A1.  
 EN 30-APR-1998.  
 PD 24-OCT-1997; 97WO-US019784.  
 PF 25-OCT-1996; 96US-00738262.  
 XX (DAND ) DANA FARBER CANCER INST INC.  
 PA Gong J, Kufe D;  
 PI WPI; 1998-261182/23.  
 XX Use of tandem repeat segment of mucin - as target for modulating effects  
 PT of cytotoxic T lymphocytes in the treatment of e.g. autoimmune disease,  
 PT transplant rejection or tumours.  
 XX

PS Disclosure; Page 28; 45pp; English.

CC The tandem repeat segment of a mucin such as DF3/MUC1 can be used in an

CC in vivo method for decreasing or preventing the deleterious effects of

CC cytotoxic T lymphocytes (CTLs) in an animal. It can be used to treat

CC disorders characterised by activation of CTLs, such as autoimmune

CC diseases (e.g. diabetes, lupus and multiple sclerosis) and conditions

CC characterised by CTL induced granuloma formation, e.g. tuberculosis, and

CC sarcoidosis, leprosy, Crohn's disease, hypersensitivity pneumonitis, and

CC primary biliary cirrhosis. The methods can also be used to treat

CC transplant rejection and malignant disorders of CTLs, e.g. activated T

CC cell leukaemia/lymphoma

XX

SQ Sequence 40 AA;

Query Match 100.0%; Score 127; DB 2; Length 40;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23

Db 5 PDTRPAPGSTAPPAGHGVTSAPDT 27

RESULT 16

AAW72703

ID AAW72703 standard; peptide; 40 AA.

XX

AC AAW72703;

XX

DT 11-JAN-1999 (first entry)

XX

DE Human mucin peptide.

XX

XX Mucin, muc-1; human; cancer; infectious disease; vaccine; diagnosis;

KW multiple tandem repeat; pancreatic cancer; breast cancer; colon cancer.

XX

OS Homo sapiens.

XX

PN US5827666-A.

XX

PD 27-OCT-1998.

XX

PF 10-AUG-1994; 94US-00288059.

XX

PR 30-JUL-1993; 93US-00099354.

XX

PA (UYPI-) UNIV PITTSBURGH.

XX

PI Montelaro RC, Fontenot JD, Finn OJ;

XX

DR WPI; 1998-593988/50.

XX

PT Assay for cancer antibodies - using synthetic peptide comprising multiple

PT tandem repeats of muc-1.

XX

PS Disclosure; Col 15; 45pp; English.

XX

CC An assay has been developed for antibodies to pancreatic, breast or colon

CC cancer in a sample. The assay comprises contacting the sample with a

CC synthetic muc-1 peptide that comprises at least two 20 amino acid tandem

CC repeats of muc-1 and is capable of attaining native conformation in the

CC absence of glycosylation, and detecting any peptide-antibody complex

CC formation. The assay can be used in the diagnosis of e.g. pancreatic,

CC breast or colon cancer. The present sequence represents a human mucin

CC peptide from the present invention

XX

SQ Sequence 40 AA;

Query Match 100.0%; Score 127; DB 2; Length 40;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23

Db 2 PDTRPAPGSTAPPAGHGVTSAPDT 24

RESULT 17

ABP56039

ID ABP56039 standard; peptide; 40 AA.

XX

AC ABP56039;

XX

DT 26-FEB-2003 (first entry)

XX

DE Repeat motif 2 (RM2) SEQ ID NO:9.

XX

KW Tumour; cancer; tumorigenesis inhibition; diagnosis; tumour marker;

KW cell proliferation; MUC1; mucin 1 receptor; MUC1 receptor; MGFR;

KW MUC1 growth factor receptor; MUC1 growth factor receptor modulators;

KW PSMGFR; cytostatic.

XX

OS Homo sapiens.

OS Synthetic.

XX

PN WO200256022-A2.

XX

PD 18-JUL-2002.

XX

PF 27-NOV-2001; 2001WO-US044782.

XX

PR 27-NOV-2000; 2000US-0253361P.

PR 13-DEC-2000; 2000US-0255370P.

PR 15-DEC-2000; 2000US-0256027P.

PR 22-DEC-2000; 2000US-0258157P.

PR 03-JAN-2001; 2001US-0259615P.

PR 05-JAN-2001; 2001US-0260186P.

PR 02-FEB-2001; 2001US-0266169P.

PR 06-FEB-2001; 2001US-0266929P.

PR 23-MAR-2001; 2001US-0278093P.

PR 07-MAY-2001; 2001US-0289444P.

PR 31-MAY-2001; 2001US-0294887P.

PR 14-JUN-2001; 2001US-0298272P.

XX

PA (MINE-) MINERVA BIOTECHNOLOGIES CORP.

XX

PI Bamdad CC, Bamdad RS;

XX

DR WPI; 2002-599674/64.

XX

PT Treating or diagnosing cell proliferation, particularly a cancer

PT characterized by aberrant expression of a MUC1 receptor (e.g. breast or

PT prostate cancer) comprises administering agents that modulate a MUC1

PT growth factor receptor.

XX

PS Example 1a; Page 54; 129pp; English.

XX

CC The present invention describes a method for treating a subject to reduce

CC the risk of or progression of cancer by administering an agent for: (a)

CC inhibiting interaction of an activating ligand with a portion of a cell

CC surface receptor that interacts with the activating ligand to promote

CC cell proliferation; or (b) preventative clustering of portions of cell

CC surface receptors that interact with an activating ligand. Also

CC described: (1) compositions comprising: (a) a portion of a shed cell

CC surface receptor interchain binding region; and (b) a signalling entity

CC immobilised relative to or adapted to be immobilised relative to the

CC portion; and (2) a peptide species comprising a fragment of a sequence

CC that corresponds to that portion of a cell surface receptor that

CC interacts with an activating ligand such as a growth factor to promote

CC cell proliferation, the portion being detached from any cell, and an

CC affinity tag. The compositions have cytostatic activity and can be used

CC as MUC1 growth factor receptor (MGFR) modulators. The method is useful

CC for treating or diagnosing cell proliferation, particularly cancer of the

CC breast, prostate, lung, ovary, colorectal or brain. In particular, the

CC cancer is characterised by aberrant expression of MUC1 receptor. The





KW glycolipopeptide; cytostatic; antibacterial; virucide; antiparasitic;  
KW vaccine; gene therapy; immunotherapeutic; cancer; cross-reactive epitope;  
KW immune response.

XX Synthetic.

OS WO2003089574-A2.

PN 30-OCT-2003.

PD 09-APR-2003; 2003WO-US010750.

PF 15-APR-2002; 2002US-0372105P.

PR 06-MAY-2002; 2002US-0377595P.

XX (BIOM-) BIOMIRA INC.

PA Koganty RR, Jiaog Z, Yalamati D, Gandhi S, Budzynski W;

PI Krantz MJ, Longenecker BM;

XX WPI; 2003-865440/80.

DR New glycolipopeptide comprising a disease-associated epitope or a  
XX lipidated interior amino acid, useful as an immunotherapeutic, as a  
PT vaccine against cancers and pathogens (e.g. virus or bacteria), or as  
PT diagnostic reagents.

XX Disclosure; SEQ ID NO 2; 167pp; English.

XX The invention relates to a novel non-naturally occurring glycolipopeptide  
CC comprising at least 5 amino acids, at least one amino acid being a  
CC glycosylated amino acid and at least one amino acid being a lipidated  
CC amino acid, where at least one lipidated amino acid is an interior amino  
CC acid, the glycolipopeptide comprising at least one disease-associated  
CC epitope. A peptide of the invention has cytostatic, antibacterial,  
CC virucide, and antiparasitic activity, and may have a use as a vaccine,  
CC and in gene therapy. The glycolipopeptide is useful as an  
CC immunotherapeutic or as a vaccine against cancers and pathogens (e.g.  
CC microbes, toxins, parasites or viruses) presenting cross-reactive  
CC epitopes, or as diagnostic reagents. The lipidated amino acid is useful  
CC for specific modulation of immune responses to an antigen. The present  
CC sequence is used in the exemplification of the invention.

XX Sequence 43 AA;

Query Match 100.0%; Score 127; DB 7; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.8e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 4 PDTRPAPGSTAPPAGHGVTSAPDT 26

RESULT 21  
ABP56035  
ID ABP56035 standard; peptide; 46 AA.

XX ABP56035;

DT 26-FEB-2003 (first entry)

XX Histidine-tagged repeat motif 2 (His-RM2) SEQ ID NO:5.

KW Tumour; cancer; tumorigenesis inhibition; diagnosis; tumour marker;  
KW cell proliferation; MUC1; mucin 1 receptor; MUC1 receptor; MGFR;  
KW MUC1 growth factor receptor; MUC1 growth factor receptor modulators;  
KW PSMGFR; cytostatic.

XX Homo sapiens.  
OS Synthetic.

XX WO200256022-A2.

XX 18-JUL-2002.  
XX 27-NOV-2001; 2001WO-US044782.  
XX 27-NOV-2000; 2000US-0253361P.  
PR 13-DEC-2000; 2000US-0255370P.  
PR 15-DEC-2000; 2000US-0256027P.  
PR 22-DEC-2000; 2000US-0258157P.  
PR 03-JAN-2001; 2001US-0259615P.  
PR 05-JAN-2001; 2001US-0260186P.  
PR 02-FEB-2001; 2001US-0266169P.  
PR 06-FEB-2001; 2001US-0266929P.  
PR 23-MAR-2001; 2001US-0278093P.  
PR 07-MAY-2001; 2001US-0289444P.  
PR 31-MAY-2001; 2001US-0294887P.  
PR 14-JUN-2001; 2001US-0298272P.

XX (MINE-) MINERVA BIOTECHNOLOGIES CORP.

XX Bamdad CC, Bamdad RS;

PI WPI; 2002-599574/64.

XX Treating or diagnosing cell proliferation, particularly a cancer  
PT characterized by aberrant expression of a MUC1 receptor (e.g. breast or  
PT prostate cancer) comprises administering agents that modulate a MUC1  
PT growth factor receptor.

XX Example 1a; Page 53; 129pp; English.

XX The present invention describes a method for treating a subject to reduce  
CC the risk of or progression of cancer by administering an agent for: (a)  
CC inhibiting interaction of an activating ligand with a portion of a cell  
CC surface receptor that interacts with the activating ligand to promote  
CC cell proliferation; or (b) preventative clustering of portions of cell  
CC surface receptors that interact with an activating ligand. Also  
CC described: (1) compositions comprising: (a) a portion of a shed cell  
CC surface receptor interchain binding region; and (b) a signalling entity  
CC immobilised relative to or adapted to be immobilised relative to the  
CC portion; and (2) a peptide species comprising a fragment of a sequence  
CC that corresponds to that portion of a cell surface receptor that  
CC interacts with an activating ligand such as a growth factor to promote  
CC cell proliferation, the portion being detached from any cell, and an  
CC affinity tag. The compositions have cytostatic activity and can be used  
CC as MUC1 growth factor receptor (MGFR) modulators. The method is useful  
CC for treating or diagnosing cell proliferation, particularly cancer of the  
CC breast, prostate, lung, ovary, colorectal or brain. In particular, the  
CC cancer is characterised by aberrant expression of MUC1 receptor. The  
CC present sequence represents a histidine-tagged repeat motif 2 (His-RM2),  
CC which is used in the exemplification of the present invention

XX Sequence 46 AA;

Query Match 100.0%; Score 127; DB 5; Length 46;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRFAPGSTAPPAGHGVTSAPDT 23  
Db 1 PDTRFAPGSTAPPAGHGVTSAPDT 23

RESULT 22  
AAE33936  
ID AAE33936 standard; peptide; 46 AA.

XX AAE33936;

XX 02-MAY-2003 (first entry)

XX MUC 1 peptide #1.



CC A cysteine is added to the N-terminal of this peptide to facilitate the  
CC conjugation with protein carriers. The immunogenic protein is a keyhole  
CC limpet haemocyanin (KLH) or its derivative. The vaccine can be used to  
CC induce an immune response in patients suffering from a cancer of the type  
CC where the cancer cells have mucin on their surface, e.g. breast cancer,  
CC prostate cancer, lung cancer, colon cancer or pancreas cancer  
XX  
SQ Sequence 51 AA;

Query Match 100.0%; Score 127; DB 2; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.2e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 6 PDTRPAPGSTAPPAGHGVTSAPDT 28

RESULT 25  
ABB76181  
ID ABB76181 standard; protein; 100 AA.

XX AC ABB76181;

DT 22-JUL-2002 (first entry)

DE Synthetic MUC1 repeat polypeptide.

XX MUC1; tumour; antigen; cancer; immunostimulant; immunotherapy;  
KW antitumour; vaccine; human; BCG.

OS Homo sapiens.  
OS Synthetic.

PN WO200226819-A2.

PD 04-APR-2002.

PF 26-SEP-2001; 2001WO-US030346.

PR 26-SEP-2000; 2000US-0235455P.

XX (WILL-) WILLIAMS HOSPITAL ROGER.  
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Chung MA, Sharma S, Chang HR, O'donnell MA;

XX WPI; 2002-405046/43.

XX Recombinant bacteria, useful for preventing and treating cancer,  
PT especially breast cancer, are engineered to secrete cytokines and express  
PT tumor antigens.

XX Example H; Page 35; 54pp; English.

CC The present sequence is a polypeptide comprising 5 tandem repeats of  
CC MUC1, a candidate tumour associated antigen for breast cancer. The  
CC polypeptide was compared with recombinant vaccine BCG-hIL2MUC1 in the  
CC vaccination of hu-PBL-SCID mice against breast cancer. BCG-hIL2MUC1 is a  
CC recombinant Mycobacterium bovis BCG bacterium that expresses a truncated  
CC form of MUC1 protein with 22 tandem repeats while simultaneously  
CC secreting human interleukin-2 (IL-2). Mice immunised with BCG-IL2MUC1 had  
CC a greater rate of survival (60%) over a 15 week period than did those  
CC immunised with phosphate buffered saline (0%), BCG-261 (0%) or the  
CC present MUC1 polypeptide (23%). Mean tumour size was significantly  
CC smaller in the mice immunised with BCG-IL2MUC1. BCG-IL2MUC1 is an example  
CC of a recombinant bacterium of the invention that has been engineered to  
CC secrete a cytokine, i.e. IL-2, and to express a tumour antigen, i.e.  
CC MUC1. The recombinant bacteria are used in claimed methods of inhibiting  
CC the growth or proliferation, or inducing the killing, of a tumour in a  
CC subject, and of treating cancer, especially prostate cancer, colon  
CC cancer, lung cancer, pancreatic cancer, ovarian cancer and particularly  
CC breast cancer (all claimed), and in a claimed vaccine for immunising a

CC subject against a neoplastic disease  
XX Sequence 100 AA;

Query Match 100.0%; Score 127; DB 5; Length 100;  
Best Local Similarity 100.0%; Pred. No. 4.3e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 6 PDTRPAPGSTAPPAGHGVTSAPDT 28

RESULT 26  
AAR68022  
ID AAR68022 standard; peptide; 105 AA.

XX AC AAR68022;

DT 25-MAR-2003 (revised)

DT 05-SEP-1995 (first entry)

DE Mucin peptide p105.

XX Mucin; multiple tandem repeat; vaccine; pancreas cancer; breast cancer;  
KW ovary cancer; colon cancer; HIV.

OS Synthetic.

XX Key Location/Qualifiers  
FH Peptide 6..25  
FT /label= Repeat\_1  
FT Peptide 26..45  
FT /label= Repeat\_2  
FT Peptide 46..65  
FT /label= Repeat\_3  
FT Peptide 66..85  
FT /label= Repeat\_4  
FT Peptide 86..105  
FT /label= Repeat\_5

XX WO9503825-A1.

XX 09-FEB-1995.

XX 29-JUL-1994; 94WO-US008477.

XX 30-JUL-1993; 93US-00099354.

XX (FINN/) FINN O J.

XX (FONT/) FONTENOT J D.

XX (MONT/) MONTELLARO R C.

XX Finn OJ, Fontenot JD, Montelaro RC;

XX WPI; 1995-082033/11.

XX Synthetic multiple tandem repeat mucin-1 peptides and analogues - have  
PT native conformation in the absence of glycosylation and are linked to  
PT epitopes; for vaccines and tests of cancer, viruses and bacteria.

XX Disclosure; Page 54; 125pp; English.

XX A synthetic peptide such as p105 (AAR68022) includes 5 tandem repeats of  
CC the mucin peptide given in AAR68004. The DTR motif, located between the  
CC first 2 prolines of each repeat, is the target of an anti-mucin immune  
CC response, and can be substituted by a sequence from a virus, tumor  
CC antigen or autoantigen. Drugs based on 9-amino acid portions (AAR68007-  
CC 21) of the mucin peptide were developed. (Updated on 25-MAR-2003 to  
CC correct PN field.)

XX Sequence 105 AA;

Query Match 100.0%; Score 127; DB 2; Length 105;  
Best Local Similarity 100.0%; Pred. No. 4.5e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 6 PDTRPAPGSTAPPAGHGVTSAPDT 28

RESULT 27  
AAW72697  
ID AAW72697 standard; peptide; 105 AA.  
XX AC AAW72697;  
XX DT 11-JAN-1999 (first entry)  
XX DE Synthetic muc-1 peptide.

KW Mucin; muc-1; human; cancer; infectious disease; vaccine; diagnosis;  
KW multiple tandem repeat; pancreatic cancer; breast cancer; colon cancer.

OS Homo sapiens.  
OS Synthetic.

PN US5827666-A.

PD 27-OCT-1998.

XX 10-AUG-1994; 94US-00288059.

XX 30-JUL-1993; 93US-00099354.

XX (UXPI-) UNIV PITTSBURGH.

PI Montelaro RC, Fontenot JD, Finn OJ;

XX WPI; 1998-593988/50.

DR Assay for cancer antibodies - using synthetic peptide comprising multiple tandem repeats of muc-1.

XX Claim 8; Col 60; 45pp; English.

XX An assay has been developed for antibodies to pancreatic, breast or colon cancer in a sample. The assay comprises contacting the sample with a synthetic muc-1 peptide that comprises at least two 20 amino acid tandem repeats of muc-1 and is capable of attaining native conformation in the absence of glycosylation, and detecting any peptide-antibody complex formation. The present sequence represents a specifically claimed synthetic muc-1 peptide. The assay can be used in the diagnosis of e.g. pancreatic, breast or colon cancer

XX Sequence 105 AA;

Query Match 100.0%; Score 127; DB 2; Length 105;  
Best Local Similarity 100.0%; Pred. No. 4.5e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 6 PDTRPAPGSTAPPAGHGVTSAPDT 28

RESULT 28  
AAW72665  
ID AAW72665 standard; peptide; 216 AA.

XX AC AAW72665;

XX 10-AUG-2000 (first entry)

XX DE MUC-1 analogue containing foreign epitopes.

XX Mucin repeat; MUC-1 analogue; vaccination; self-protein; cancer;  
KW cytotoxic T-lymphocyte immunity; breast cancer; prostate cancer;  
KW cell-associated peptide antigen; foreign epitope.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH Peptide 61..75  
FT /label= P2  
FT Peptide 136..156  
FT /label= P30  
FT /note= "q"

XX WO200020027-A2.

XX 13-APR-2000.

XX 05-OCT-1999; 99WO-DK000525.

XX 05-OCT-1998; 98DK-00001261.

XX 20-OCT-1998; 98US-0105011P.

XX (MEBI-) M & E BIOTECH AS.

XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;

XX Gautam A, Birk P, Karlsson G;

XX WPI; 2000-349917/30.

XX Inducing immune responses to weakly immunogenic, tumor associated peptide antigens for the treatment of breast and prostate cancer.

XX Example 4; Page; 220pp; English.

XX This is an immunogenized MUC-1 analogue containing foreign epitopes P2 and P30. Immunogenic analogues of MUC-1 and, e.g. human prostate specific membrane antigen (hPSM) can be used in the claimed method as an autovaccine to induce a CTL response. Subdominant CTL epitopes, antibody binding regions and cysteine residues involved in disulfide bonds are preserved in the immunogenized forms (see features table). 10 regions suitable for the insertion of foreign T helper epitopes were identified. The method is used for inducing immune responses against weakly immunogenic cell-associated peptide antigens (PA) such as those associated with cancers (self-proteins), e.g. hPSM, heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b). The method comprises effecting simultaneous presentation by antigen producing cells (APCs) of the animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1 B-cell group derived from the cell-associated PA; and (2) at least 1 first T helper cell group which is foreign to the animal. Analogues of human PSM, human Her2 and human/murine FGF8b comprising a substantial part of all known and predicted CTL and B-cell epitopes of the respective PA and including at least one foreign T helper epitope are also claimed. The method is used to treat prostate, prostate/breast or breast cancer when the PA is human PSM, FGF8b and Her2, respectively. Note: This sequence does not appear in the specification. It was made using the mucin repeat sequence (AAW92664), P2 and P30 (AAW92625-26), which appear on pages 220, 213 and 214 respectively, of the specification

XX Sequence 216 AA;

Query Match 100.0%; Score 127; DB 3; Length 216;  
Best Local Similarity 100.0%; Pred. No. 9.2e-08;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 17 PDTRPAPGSTAPPAGHGVTSAPDT 39

RESULT 29  
AAW77233

ID AAW77233 standard; protein; 508 AA.  
 AC AAW77233;  
 XX  
 DT 20-NOV-1998 (first entry)  
 DE  
 XX MiniMUC1 protein.  
 DE  
 KW MUC1; recombinant pox virus; cytotoxic T-lymphocyte; immunogen; tumour;  
 KW tumour-associated antigen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9837095-A2.  
 XX  
 PD 27-AUG-1998.  
 XX  
 PF 24-FEB-1998; 98WO-US003693.  
 XX  
 PR 24-FEB-1997; 97US-0038253P.  
 XX  
 PA (THER-) THERION BIOLOGICS CORP.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA (DAND ) DANA FARBER CANCER INST INC.  
 XX  
 PI Schlom J, Kantor J, Kufe D, Panicali D, Gritz L;  
 DR WPI; 1998-467492/40.  
 DR N-PSDB; AAV48329.  
 XX  
 PT New recombinant pox virus for tumour therapy - comprises DNA encoding an  
 PT immunogenic mini-MUC1 fragment comprising 5-25 MUC1 tandem repeat units.  
 XX  
 PS Example 1; Page 21-22; 42pp; English.  
 XX  
 CC The immunogenic mini-MUC1 fragment was created from MUC1 tandem repeat  
 CC units for inclusion in a recombinant pox virus (RPV). The RPV was used in  
 CC a pharmaceutical composition also containing an immunomodulator to  
 CC generate MUC1 specific cytotoxic T-lymphocytes. The recombinant pox virus  
 CC therefore encodes an immunogenic MUC1 fragment that does not undergo  
 CC significant genetic deletion, thereby providing an unexpectedly stable  
 CC and immunogenic pox virus. They can be used to prevent or treat tumours  
 CC expressing MUC1 tumour-associated antigens  
 XX  
 SQ Sequence 508 AA;  
 Query Match 100.0%; Score 127; DB 2; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 PDTRPAPGSGTAPPAHGVTSAPDT 23  
 Db 89 PDTRPAPGSGTAPPAHGVTSAPDT 111  
 RESULT 30  
 ABP56040  
 ID ABP56040 standard; peptide; 1255 AA.  
 XX  
 AC ABP56040;  
 XX  
 DT 26-FEB-2003 (first entry)  
 XX  
 DE MUC1 receptor (mucin 1 precursor) SEQ ID NO:10.  
 XX  
 KW Tumour; cancer; tumorigenesis inhibition; diagnosis; tumour marker;  
 KW cell proliferation; MUC1; mucin 1 receptor; MUC1 receptor; MGFR;  
 KW MUC1 growth factor receptor; MUC1 growth factor receptor modulators;  
 KW PSMGFR; cytostatic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200256022-A2.

XX 18-JUL-2002.  
 PD  
 XX 27-NOV-2001; 2001WO-US044782.  
 PF  
 XX 27-NOV-2000; 2000US-0253361P.  
 PR 13-DEC-2000; 2000US-0255370P.  
 PR 15-DEC-2000; 2000US-0256027P.  
 PR 22-DEC-2000; 2000US-0258157P.  
 PR 03-JAN-2001; 2001US-0259615P.  
 PR 05-JAN-2001; 2001US-0260186P.  
 PR 02-FEB-2001; 2001US-0266169P.  
 PR 06-FEB-2001; 2001US-0266929P.  
 PR 23-MAR-2001; 2001US-0278093P.  
 PR 07-MAY-2001; 2001US-0289444P.  
 PR 31-MAY-2001; 2001US-0294887P.  
 PR 14-JUN-2001; 2001US-0298272P.  
 XX  
 PA (MINE-) MINERVA BIOTECHNOLOGIES CORP.  
 PI Bamdad CC, Bamdad RS;  
 XX WPI; 2002-599674/64.  
 DR  
 XX Treating or diagnosing cell proliferation, particularly a cancer  
 PT characterized by aberrant expression of a MUC1 receptor (e.g. breast or  
 PT prostate cancer) comprises administering agents that modulate a MUC1  
 PT growth factor receptor.  
 XX  
 PS Example 1a; Page 54-55; 129pp; English.  
 XX  
 CC The present invention describes a method for treating a subject to reduce  
 CC the risk of or progression of cancer by administering an agent for: (a)  
 CC inhibiting interaction of an activating ligand with a portion of a cell  
 CC surface receptor that interacts with the activating ligand to promote  
 CC cell proliferation; or (b) preventative clustering of portions of cell  
 CC surface receptors that interact with an activating ligand. Also  
 CC described: (1) compositions comprising: (a) a portion of a shed cell  
 CC surface receptor interchain binding region; and (b) a signalling entity  
 CC immobilised relative to or adapted to be immobilised relative to the  
 CC portion; and (2) a peptide species comprising a fragment of a sequence  
 CC that corresponds to that portion of a cell surface receptor that  
 CC interacts with an activating ligand such as a growth factor to promote  
 CC cell proliferation, the portion being detached from any cell, and an  
 CC affinity tag. The compositions have cytostatic activity and can be used  
 CC as MUC1 growth factor receptor (MGFR) modulators. The method is useful  
 CC for treating or diagnosing cell proliferation, particularly cancer of the  
 CC breast, prostate, lung, ovary, colorectal or brain. In particular, the  
 CC cancer is characterised by aberrant expression of MUC1 receptor. The  
 CC present sequence represents a MUC1 receptor (mucin 1 precursor), which is  
 CC used in the exemplification of the present invention  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 100.0%; Score 127; DB 5; Length 1255;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-07;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 PDTRPAPGSGTAPPAHGVTSAPDT 23  
 Db 142 PDTRPAPGSGTAPPAHGVTSAPDT 164  
 RESULT 31  
 ABB82568  
 ID ABB82568 standard; protein; 1255 AA.  
 XX  
 AC ABB82568;  
 XX  
 DT 04-FEB-2003 (first entry)  
 XX  
 DE MUC1/REP transmembrane protein.  
 XX



KW MUC1; mucin; cell proliferation; cell growth; cell death; cytostatic;  
KW gene therapy; cancer; tumour; MUC1/REP; transmembrane.  
XX Homo sapiens.  
OS  
XX WO200278598-A2.  
PN  
XX 10-OCT-2002.  
PD  
XX 26-MAR-2002; 2002WO-IL000255.  
PF  
XX 29-MAR-2001; 2001US-0279408P.  
PR  
XX (UWRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
PA  
XX Wreschner DH, Smorodinsky NI, Yoeli-Lerner M;  
PI WPI; 2003-040613/03.  
DR  
XX Inhibiting proliferation or growth of tumor cells and/or inducing cell  
PT death, comprises administering a ligand which binds to an epitope in the  
PT extracellular region of a transmembrane isoform of MUC1 protein.  
PT  
XX Disclosure; Fig 5; 42pp; English.  
PS  
XX The invention relates to inhibiting mammalian cell proliferation or cell  
CC growth and/or inducing mammalian cell death and involves administering to  
CC a subject an amount of a ligand, which specifically binds to an epitope  
CC in the extracellular region of a transmembrane isoform of MUC1 protein to  
CC selectively inhibit mammalian cell proliferation or cell growth or induce  
CC mammalian cell death. The method is useful in inhibiting proliferation or  
CC growth of tumour cells and/or inducing death in epithelial, colon,  
CC breast, lungs or ovarian tumour cells or other cells expressing MUC1  
CC proteins, and in treating patients with a disease involving pathological  
CC proliferation of cells. The present sequence represents an isoform of the  
CC MUC1 protein, MUC1/REP, a transmembrane protein  
XX  
SQ Sequence 1255 AA;

Query Match 100.0%; Score 127; DB 6; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 5.4e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 142 PDTRPAPGSTAPPAGHGVTSAPDT 164  
  
RESULT 32  
ABR47537  
ID ABR47537 standard; protein; 1255 AA.  
XX  
AC ABR47537;  
XX  
DT 12-JUN-2003 (first entry)  
XX  
DE Breast cancer associated protein; sequence SEQ ID NO:311.  
XX  
KW Human; breast cancer; cytostatic; gene therapy.  
KW  
XX Homo sapiens.  
OS  
XX WO2003004989-A2.  
PN  
XX 16-JAN-2003.  
PD  
XX 21-JUN-2002; 2002WO-US019669.  
PF  
XX 21-JUN-2001; 2001US-0299887P.  
PR  
XX 27-JUN-2001; 2001US-0301572P.  
PR  
XX 18-JUL-2001; 2001US-0306501P.  
PR  
XX 25-SEP-2001; 2001US-0325002P.  
PR  
XX 05-MAR-2002; 2002US-0362585P.

PR 14-MAY-2002; 2002US-0380391P.  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kamatkar S;  
PI Mertens M, Monahan JB, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;  
PI Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;  
XX WPI; 2003-210381/20.  
DR N-PSDB; ACC50234.  
DR  
XX Breast cancer diagnosis or treatment by comparing the level of expression  
PT of a marker in a patient sample with that in the control non-breast  
PT cancer sample.  
XX  
PS Claim 1; SEQ ID NO 311; 128pp; English.  
XX  
CC The present invention describes a method for assessing whether a patient  
CC is afflicted with breast cancer. The method comprises comparing the level  
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and  
CC ABR47386 to ABR47632) in a patient sample and the normal level of  
CC expression of the marker in a control non-breast cancer sample, where a  
CC significant increase in the level of expression of the marker in the  
CC patient sample and the normal level is an indication that the patient is  
CC afflicted with breast cancer. The breast cancer associated sequences from  
CC the present invention have cytostatic activities and can be used in gene  
CC therapy. The method is useful for diagnosing and treating breast cancer.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1255 AA;

Query Match 100.0%; Score 127; DB 6; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 5.4e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 142 PDTRPAPGSTAPPAGHGVTSAPDT 164  
  
RESULT 33  
ABR92124  
ID ABR92124 standard; protein; 1255 AA.  
XX  
AC ABR92124;  
XX  
DT 10-SEP-2003 (first entry)  
XX  
DE Human cervical cancer cell marker protein SEQ ID NO:158.  
XX  
KW Human; cervical cancer; cervical cancer marker; cancer therapy;  
KW detection; gene therapy; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO2002101075-A2.  
XX  
PD 19-DEC-2002.  
XX  
PF 12-JUN-2002; 2002WO-US018638.  
XX  
PR 13-JUN-2001; 2001US-0298155P.  
PR 13-JUN-2001; 2001US-0298159P.  
PR 14-NOV-2001; 2001US-0335936P.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Schlegel R, Chen Y, Zhao X, Monahan JE, Kamatkar S;  
PI Gannavarapu M, Glatt K, Hoersch S;  
XX WPI; 2003-156967/15.  
DR

DR N-PSDB; ACF12907.

XX New isolated nucleic acid molecule useful for detecting, characterizing, preventing and treating human cervical cancers, in various prognostic and diagnostic assays, in pharmacogenomics and in monitoring clinical trials.

PT

PT

XX

PS Claim 4; Page 312-315; 386pp; English.

XX

CC ACF12828 to ACF12947 encode the human cervical cancer marker proteins (I) given in ABR92047 to ABR92164. A higher level of expression of (I) than normal indicates the presence of cervical cancer. Also described: (1) a vector (II) containing (I); (2) a host cell (III) containing (I); and (3) assessing (M) whether a patient is afflicted with cervical cancer, comprising comparing the level of expression of a marker in a patient's sample, and the normal level of expression of the marker in a control non-cervical cancer sample, where a significant increase in the level of expression of the marker in the patient's sample relative to that in the control sample is an indication that the patient is afflicted with cervical cancer. (I) has cytostatic activity, and can be used in gene therapy and in vaccines. (I) is useful in detecting, characterizing, preventing and treating human cervical cancers. (I) may also be used in various prognostic and diagnostic assays, pharmacogenomics and in monitoring clinical trials

XX

SQ Sequence 1255 AA;

Query Match 100.0%; Score 127; DB 6; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 5.4e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGCTAPPAGHGVTSAPDT 23  
|||||

Db 142 PDTRPAPGCTAPPAGHGVTSAPDT 164  
|||||

RESULT 34

ADD45111

ID ADD45111 standard; protein; 1255 AA.

XX AC ADD45111;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human Protein Q16615, SEQ ID NO 10544.

XX

XW Human; pain; neuronal tissue; gene therapy;

XW spinal segmental nerve injury; chronic constriction injury; CCI;

XW spared nerve injury; SNI; Chung.

XX

OS Homo sapiens.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX

PA (GEHO ) GEN HOSPITAL CORP.

PA (FARB ) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

DR WPI; 2003-268312/26.

DR GENBANK; Q16615.

XX

PT New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

XX

ES Claim 1; Page; 1017pp; English.

XX

CC The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 1255 AA;

Query Match 100.0%; Score 127; DB 7; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 5.4e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGCTAPPAGHGVTSAPDT 23  
|||||

Db 142 PDTRPAPGCTAPPAGHGVTSAPDT 164  
|||||

RESULT 35

ADE54622

ID ADE54622 standard; protein; 1255 AA.

XX AC ADE54622;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human Protein Q16615, SEQ ID NO 427.

XX

XW Human; pain; neuronal tissue; gene therapy;

XW spinal segmental nerve injury; chronic constriction injury; CCI;

XW spared nerve injury; SNI; Chung.

XX

OS Homo sapiens.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX

PA (GEHO ) GEN HOSPITAL CORP.

PA (FARB ) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

DR WPI; 2003-268312/26.

DR GENBANK; Q16615.

XX New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX Claim 1; Page; 1017pp; English.  
 XX The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 1255 AA;

Query Match 100.0%; Score 127; DB 7; Length 1255;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-07;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
 Db |||||  
 142 PDTRPAPGSTAPPAGHGVTSAPDT 164

RESULT 36  
 AAY71021  
 ID AAY71021 standard; protein; 173 AA.  
 AC AAY71021;  
 XX 29-AUG-2000 (first entry)  
 DE Human Mucin 1 (MUC-1) protein fragment #2.  
 KW Human; Mucin 1; tumour; pMRS30 expression vector; anti-tumour;  
 KW therapy; immune response; cytostatic; vaccine.  
 OS Homo sapiens.  
 OS WO200025827-A2.  
 PN 11-MAY-2000.  
 PD 18-OCT-1999; 99WO-EP007874.  
 XX 30-OCT-1998; 98IT-MI02330.  
 XX (MENA ) MENARINI RICERCHE SPA.  
 PI Parente D, Di Massimo AM, De Santis R;  
 XX WPI; 2000-365410/31.

DR N-PSDB; AAD00385.  
 XX Composition containing one or more DNA molecules encoding fragments of a  
 PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-  
 PT tumor therapy.  
 XX Claim 16; Fig 2; 56pp; English.  
 XX The present sequence is a fragment of human Mucin 1 (MUC-1), an antigenic  
 CC protein overexpressed in tumour cells. The sequence was obtained from  
 CC B720 tumour cells. The corresponding DNA sequence is cloned into a pMRS30  
 CC expression vector and used in pharmaceutical composition e.g. vaccine for  
 CC inducing an antigen-specific anti-tumour immune response. Composition  
 CC containing this DNA molecule is useful in anti-tumour therapy of patients  
 CC affected with tumours characterised by high MUC-1 expression  
 XX Sequence 173 AA;

Query Match 96.1%; Score 122; DB 3; Length 173;  
 Best Local Similarity 100.0%; Pred. No. 3e-07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22  
 Db |||||  
 99 PDTRPAPGSTAPPAGHGVTSAPD 120

RESULT 37  
 AAY71027  
 ID AAY71027 standard; protein; 295 AA.  
 XX AAY71027;  
 AC AAY71027;  
 XX 12-SEP-2003 (revised)  
 DT 29-AUG-2000 (first entry)

DE Ubiquitin-E. coli LacI-human Mucin 1 fusion protein #2.  
 KW Ubiquitin; LacI; beta-galactosidase; fusion protein; human; Mucin 1;  
 KW MUC-1; tumour; pMRS30 expression vector; anti-tumour; therapy;  
 KW immune response; cytostatic; vaccine.  
 OS Homo sapiens.  
 OS Escherichia coli.  
 OS Chimeric.

Key Location/Qualifiers  
 Region 1..123  
 /label= UBILacI\_protein  
 /note= "contains ubiquitin sequence fused to a portion of  
 E. coli LacI"  
 Region 124..295  
 /label= Human\_MUC-1\_fragment  
 WO200025827-A2.  
 11-MAY-2000.  
 18-OCT-1999; 99WO-EP007874.  
 30-OCT-1998; 98IT-MI02330.  
 (MENA ) MENARINI RICERCHE SPA.  
 Parente D, Di Massimo AM, De Santis R;  
 WPI; 2000-365410/31.  
 N-PSDB; AAD00391.

Composition containing one or more DNA molecules encoding fragments of a  
 PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-  
 PT tumor therapy.  
 XX

PS Claim 18; Fig 8; 56pp; English.

XX The present sequence is a fusion protein consisting of human Mucin 1 (MUC  
CC -1) fragment fused to Ubilaci sequence at the N-terminus. The Ubilaci  
CC sequence consists of ubiquitin from MCF7 cell line and a portion of E.  
CC coli beta-galactosidase (Laci). MUC-1 is an antigenic protein  
CC overexpressed in tumour cells. The corresponding DNA sequence is cloned  
CC into a pMRS30 expression vector and used in pharmaceutical composition  
CC e.g. vaccine for inducing an antigen-specific anti-tumour immune  
CC response. Composition containing this DNA molecule is useful in anti-  
CC tumour therapy of patients affected with tumours characterised by high  
CC MUC-1 expression. (Updated on 12-SEP-2003 to standardise OS field)

XX SQ Sequence 295 AA;

Query Match 96.1%; Score 122; DB 3; Length 295;  
Best Local Similarity 100.0%; Pred. No. 5.1e-07;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22  
Db 221 PDTRPAPGSTAPPAGHGVTSAPD 242

RESULT 38  
AAAY71024  
ID AAY71024 standard; protein; 455 AA.

AC AAY71024;

DT 29-AUG-2000 (first entry)

DE Human Mucin 1 (MUC-1) protein fragment #5.

KW Human; Mucin 1; MUC-1; tumour; pMRS30 expression vector; anti-tumour;  
KW therapy; immune response; cytostatic; vaccine.

XX Homo sapiens.

OS WO200025827-A2.

PN 11-MAY-2000.

PF 18-OCT-1999; 99WO-EP007874.

PR 30-OCT-1998; 98IT-MI002330.

PA (MENA ) MENARINI RICERCHE SPA.

PI Parente D, Di Massimo AM, De Santis R;

DR WPI; 2000-365410/31.

XX N-PSDB; AAC0388.

XX Composition containing one or more DNA molecules encoding fragments of a  
PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-  
PT tumor therapy.

PS Claim 16; Fig 5; 56pp; English.

XX The present sequence is a fragment of human Mucin 1 (MUC-1), an antigenic  
CC protein overexpressed in tumour cells. The sequence was obtained from  
CC BT20 tumour cells. The corresponding DNA sequence is cloned into a pMRS30  
CC expression vector and used in pharmaceutical composition e.g. vaccine for  
CC inducing an antigen-specific anti-tumour immune response. Composition  
CC containing this DNA molecule is useful in anti-tumour therapy of patients  
CC affected with tumours characterised by high MUC-1 expression

XX SQ Sequence 455 AA;

Query Match 96.1%; Score 122; DB 3; Length 455;  
Best Local Similarity 100.0%; Pred. No. 7.9e-07;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22  
Db 122 PDTRPAPGSTAPPAGHGVTSAPD 143

RESULT 39

AAE09508

ID AAE09508 standard; protein; 473 AA.

XX AAE09508;

DT 19-NOV-2001 (first entry)

DE Human mucin-1 (MUC-1) protein.

KW Human milk fat globule membrane antigen; HMFG; immunostimulant;  
KW cytostatic; cell mediated immune response; carcinoma; adenocarcinoma;  
KW breast cancer; dendritic cell; vaccine; gene therapy; mucin-1; MUC-1;  
KW immunogen.

OS Homo sapiens.

XX Key Location/Qualifiers  
FT Peptide 1..21  
FT Protein /label= Signal\_peptide  
FT Region /label= Mature\_MUC-1\_protein  
FT Domain 132..170  
FT /label= VNTR\_region  
FT /label= Transmembrane\_domain

PN WO200157058-A1.

PD 09-AUG-2001.

PF 01-FEB-2001; 2001WO-AU0000090.

PR 01-FEB-2000; 2000AU-00005369.

XX 14-JUN-2000; 2000US-00593870.

PA (AUST-) AUSTIN RES INST.

PI Mckenzie IFC, Pietersz GA, Apostolopoulos V;

DR WPI; 2001-541537/60.

PT Immunostimulant peptide, used as an anti-carcinoma vaccine, comprises a  
PT an epitope of the non-VNTR, non-leader region of a mucin.

PS Disclosure; Fig 1; 84pp; English.

XX The patent discloses peptide or polypeptides capable of eliciting an  
CC immune response, comprising an amino acid sequence corresponding to an  
CC epitope of the non-central portion of varying numbers of an amino acid  
CC motif (VNTR), non-leader region of a mucin. The peptides of the  
CC invention, fusion proteins comprising the peptide and conjugate compounds  
CC with carbohydrate polymers are used to induce a cell mediated immune  
CC response against mucin in the prevention or treatment of carcinoma,  
CC preferably adenocarcinoma, most preferably breast cancer. They are also  
CC used to pulse dendritic cell for in vivo transfer and use as a vaccine.  
CC They are also used in gene therapy. The present protein sequence is human  
CC milk fat globule membrane antigen (HMFG), mucin-1 (MUC-1)

XX SQ Sequence 473 AA;

Query Match 96.1%; Score 122; DB 4; Length 473;  
Best Local Similarity 100.0%; Pred. No. 8.2e-07;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22  
|||||

Db 140 PDTRPAPGSTAPPAHGVTSAPD 161

Search completed: May 6, 2004, 16:23:09  
Job time : 61.5641 secs

RESULT 40

AAU00573

ID AAU00573 standard; protein; 475 AA.

XX AC AAU00573;

XX DT 12-SEP-2001 (first entry)

XX DE Human MUC1 polypeptide variant.

XX KW Human; MUC1; antigenic peptide; major histocompatibility complex; MHC-I;

XX KW Glycoprotein; cytotoxic T lymphocytes; T cell response; cancer; vaccine;

XX KW cancer gene therapy; diagnosis; treatment; inflammatory disorder;

XX KW organ transplant rejection; graft versus host disease.

XX OS Homo sapiens.

XX PN WO200118035-A2.

XX PD 15-MAR-2001.

XX PF 07-SEP-2000; 2000WO-EP008761.

XX PR 08-SEP-1999; 99GB-00021242.

XX PR 10-SEP-1999; 99EP-00402237.

XX PR 03-MAR-2000; 2000US-0187215P.

XX PA (TRGE ) TRANSGENE SA.

XX PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.

XX PI Taylor-Papadimitriou J, Heukamp LC, Offringa R, Melief CJM;

XX PI Acres B, Thomas M;

XX DR WPI; 2001-235187/24.

XX PT New antigenic polypeptides of MUC-1 protein which activate cytotoxic T

XX PT lymphocyte proteins and their analogs, useful for identifying a major

XX PT histocompatibility complex class I restricted T cell response and for

XX PT diagnosing cancer.

XX PS Disclosure; Fig 12; 81pp; English.

XX CC The sequence represents a human MUC1 polypeptide variant. Derivative

XX CC antigenic peptides of MUC1 protein bind at least one major

XX CC histocompatibility complex class I (MHC-I) glycoprotein, which activates

XX CC cytotoxic T lymphocytes to induce a protective response against tumours.

XX CC Diagnosis of cancer involves determining the presence or absence in a

XX CC host cell of MHC class I restricted T cell response to a MUC1 derivative,

XX CC where the presence of the MHC class I restricted T cell response

XX CC indicates that the host has cancer. Measurement of the level of MHC class

XX CC I restricted T cell response is also useful to monitor the severity of

XX CC cancer, a larger response indicating a more severe cancer. MUC1

XX CC derivatives are useful in cancer therapy and to follow MUC1 specific

XX CC immune responses in patients during the course of disease and/or

XX CC treatment. MUC1 DNA is useful in cancer gene therapy, vaccination and

XX CC diagnosis. Compositions of the sequences are used in vaccines and

XX CC treatments against cancer or diseases caused by an immune response, such

XX CC as an inflammatory disorder, organ transplant rejection or graft versus

XX CC host disease

XX SQ Sequence 475 AA;

Query Match

Best Local Similarity 96.1%; Score 122; DB 4; Length 475;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPD 22

Db 142 PDTRPAPGSTAPPAHGVTSAPD 163



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:27:19 ; Search time 19.4615 Seconds  
(without alignments)  
61.013 Million cell updates/sec

Title: US-10-070-566-7

Perfect score: 127

Sequence: 1 PDTRPAPGSTAPPAHGVTSAPDT 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	100.0	28	US-08-488-161-9	Sequence 9, Appli
2	127	100.0	28	US-09-273-685-9	Sequence 9, Appli
3	127	100.0	28	PCT-US95-11934-9	Sequence 9, Appli
4	127	100.0	30	US-08-134-198E-13	Sequence 13, Appl
5	127	100.0	30	US-09-593-870A-47	Sequence 47, Appl
6	127	100.0	32	US-09-593-870A-46	Sequence 46, Appl
7	127	100.0	40	US-08-099-354-1	Sequence 1, Appli
8	127	100.0	40	US-08-288-059-7	Sequence 7, Appli
9	116	91.3	21	US-08-833-807-14	Sequence 14, Appl
10	116	91.3	21	US-09-223-043-14	Sequence 14, Appl
11	116	91.3	21	US-09-593-870A-23	Sequence 23, Appl
12	113	89.0	172	US-09-646-028-49	Sequence 49, Appl
13	113	89.0	177	US-09-646-028-54	Sequence 54, Appl
14	109	85.8	20	US-08-288-059-1	Sequence 1, Appli
15	109	85.8	20	US-08-288-059-32	Sequence 32, Appl
16	109	85.8	20	US-08-902-516-20	Sequence 20, Appl
17	109	85.8	20	US-09-339-944-1	Sequence 1, Appli
18	109	85.8	20	US-08-737-896-3	Sequence 3, Appli
19	109	85.8	20	US-08-134-198E-34	Sequence 34, Appl
20	109	85.8	20	US-09-847-185-20	Sequence 20, Appl
21	109	85.8	20	US-09-646-028-40	Sequence 40, Appl
22	109	85.8	20	US-09-497-232-9	Sequence 9, Appli
23	109	85.8	20	US-09-651-265-1	Sequence 1, Appli
24	109	85.8	20	US-09-000-003A-3	Sequence 3, Appli
25	109	85.8	20	US-09-601-729-140	Sequence 140, App
26	109	85.8	20	PCT-US96-09951-3	Sequence 3, Appli
27	109	85.8	25	US-08-398-059-28	Sequence 28, Appl

28	109	85.8	30	3	US-08-737-896-6	Sequence 6, Appli
29	109	85.8	30	5	PCT-US96-09951-6	Sequence 6, Appli
30	105	82.7	1867	2	US-08-479-537A-5	Sequence 5, Appli
31	105	82.7	1867	3	US-09-083-116-5	Sequence 5, Appli
32	105	82.7	1867	4	US-09-134-916A-5	Sequence 5, Appli
33	105	82.7	2035	2	US-08-479-537A-2	Sequence 2, Appli
34	105	82.7	2035	3	US-09-083-116-2	Sequence 2, Appli
35	105	82.7	2035	4	US-09-134-916A-2	Sequence 2, Appli
36	104	81.9	134	4	US-09-646-028-1	Sequence 1, Appli
37	104	81.9	137	4	US-09-646-028-2	Sequence 2, Appli
38	104	81.9	138	4	US-09-646-028-3	Sequence 3, Appli
39	104	81.9	156	4	US-09-646-028-4	Sequence 4, Appli
40	101	79.5	20	2	US-08-833-807-1	Sequence 1, Appli
41	101	79.5	20	3	US-09-223-043-1	Sequence 1, Appli
42	101	79.5	20	4	US-09-593-870A-1	Sequence 1, Appli
43	99.5	78.3	23	4	US-09-043-731-17	Sequence 17, Appl
44	96	75.6	20	4	US-09-497-232-11	Sequence 11, Appl
45	94	74.0	40	1	US-08-099-354-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-488-161-9  
; Sequence 9, Application US/08488161  
; Patent No. 5885577  
; GENERAL INFORMATION:  
; APPLICANT: Alvarez, Vernon L.  
; TITLE OF INVENTION: Antigen Binding Peptides (Abridges) From  
; TITLE OF INVENTION: Peptide Libraries  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,161  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 436  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mirock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-176  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-488-161-9

Query Match 100.0%; Score 127; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGVTSAPDT 23

Db 5 PDTRPAPGSTAPPAHGVTSAPDT 27

RESULT 2  
US-09-273-685-9  
; Sequence 9, Application US/09273685  
; Patent No. 6015561  
; GENERAL INFORMATION:  
; APPLICANT: Alvarez, Vernon L.  
; TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From  
; TITLE OF INVENTION: Peptide Libraries  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/273,685  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA: 08/488,161  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-176  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-273-685-9  
  
Query Match 100.0%; Score 127; DB 3; Length 28;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23  
Db 5 PDTRPAGSTAPPAGHGVTSAPDT 27  
  
RESULT 3  
PCT-US95-11934-9  
; Sequence 9, Application PC/TUS9511934  
; GENERAL INFORMATION:  
; APPLICANT: Cytogen Corporation  
; TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From  
; TITLE OF INVENTION: Peptide Libraries  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/11934  
FILING DATE: 20-SEP-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-196-228  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-11934-9  
  
Query Match 100.0%; Score 127; DB 5; Length 28;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23  
Db 5 PDTRPAGSTAPPAGHGVTSAPDT 27  
  
RESULT 4  
US-08-134-198E-13  
; Sequence 13, Application US/08134198E  
; Patent No. 6190885  
; GENERAL INFORMATION:  
; APPLICANT: CANCER RESEARCH FUND  
; APPLICANT: OF CONTRA COSTA  
; APPLICANT: PETERSON, JERRY A.  
; APPLICANT: LARocca, DAVID J.  
; TITLE OF INVENTION: FUSION PROTEIN CONTAINING HMFG  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder & Poplawski  
; STREET: 444 South Flower Street, Suite 1900  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0,  
; SOFTWARE: Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/134,198E  
; FILING DATE: October 8, 1993  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Amzel, Viviana  
; REGISTRATION NUMBER: 30,930  
; REFERENCE/DOCKET NUMBER: P66 38208 (CRFC-003C)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 622-7700  
; TELEFAX: (213) 489-4210  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-134-198E-13

Query Match 100.0%; Score 127; DB 3; Length 30;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23  
Db 6 PDTRPAGSTAPPAGHGVTSAPDT 28

RESULT 5

US-09-593-870A-47  
Sequence 47, Application US/09593870A  
Patent No. 6548643  
GENERAL INFORMATION:  
APPLICANT: McKenzie, Ian F.C.  
APPLICANT: Apostolopoulos, Vasso  
APPLICANT: Pietersz, Geoff Allan  
TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their  
FILE REFERENCE: Use in Immunotherapy  
CURRENT APPLICATION NUMBER: US/09/593,870A  
CURRENT FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: 09/223,043  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 47  
LENGTH: 30  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-593-870A-47

Query Match 100.0%; Score 127; DB 4; Length 30;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23  
Db 1 PDTRPAGSTAPPAGHGVTSAPDT 23

RESULT 6

US-09-593-870A-46  
Sequence 46, Application US/09593870A  
Patent No. 6548643  
GENERAL INFORMATION:  
APPLICANT: McKenzie, Ian F.C.  
APPLICANT: Apostolopoulos, Vasso  
APPLICANT: Pietersz, Geoff Allan  
TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their  
FILE REFERENCE: Use in Immunotherapy  
CURRENT APPLICATION NUMBER: US/09/593,870A  
CURRENT FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: 09/223,043  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 46  
LENGTH: 32  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-593-870A-46

Query Match 100.0%; Score 127; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 3e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23  
Db 1 PDTRPAGSTAPPAGHGVTSAPDT 23

Db 1 PDTRPAGSTAPPAGHGVTSAPDT 23

RESULT 7

US-08-099-354-1  
Sequence 1, Application US/08099354  
Patent No. 5744144  
GENERAL INFORMATION:  
APPLICANT: FINN, OLIVERA J.  
APPLICANT: FONTENOT, J. D.  
APPLICANT: MONTELARO, RONALD C.  
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN  
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/099,354  
FILING DATE: 30-JUL-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SIRILLA, GEORGE M.  
REGISTRATION NUMBER: 18221  
REFERENCE/DOCKET NUMBER: 6137/202246  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3536  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-099-354-1

Query Match 100.0%; Score 127; DB 1; Length 40;  
Best Local Similarity 100.0%; Pred. No. 3.8e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23  
Db 2 PDTRPAGSTAPPAGHGVTSAPDT 24

RESULT 8

US-08-288-059-7  
Sequence 7, Application US/08288059  
Patent No. 5827666  
GENERAL INFORMATION:  
APPLICANT: FINN, OLIVERA J.  
APPLICANT: FONTENOT, J. D.  
APPLICANT: MONTELARO, RONALD C.  
TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN  
TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA

```

;
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,059
; FILING DATE: 08-AUG-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: CHAPIN, MARLANA K.
; REGISTRATION NUMBER: 35,843
; REFERENCE/DOCKET NUMBER: 61137/205204
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-288-059-7

Query Match 100.3%; Score 127; DB 2; Length 40;
Best Local Similarity 100.0%; Pred. No. 3.8e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAGHGVTSAPDT 23
Db 2 PDTRPAPGSTAPPAGHGVTSAPDT 24

RESULT 9
US-08-833-807-14
; Sequence 14, Application US/088333807
; Patent No. 5989552
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann Dorfman Herrell and Skillman
; STREET: Suite 720, 1601 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,807
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/833,807
; FILING DATE:
; APPLICATION NUMBER: AU PM3223
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215)563-4100
; TELEFAX: (215)563-4044
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-223-043-14

Query Match 91.3%; Score 116; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.1e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAGHGVTSAP 21
Db 1 PDTRPAPGSTAPPAGHGVTSAP 21

```

```

;
; TELEPHONE: (215)563-4100
; TELEFAX: (215)563-4044
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-833-807-14

Query Match 91.3%; Score 116; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.1e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAGHGVTSAP 21
Db 1 PDTRPAPGSTAPPAGHGVTSAP 21

RESULT 10
US-09-223-043-14
; Sequence 14, Application US/09223043
; Patent No. 6177256
; GENERAL INFORMATION:
; APPLICANT: McKenzie, Ian F.C.
; APPLICANT: Apostolopoulos, Vasso
; APPLICANT: Pietersz, Geoff A.
; TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND
; TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann Dorfman Herrell and Skillman
; STREET: Suite 720, 1601 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/223,043
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/833,807
; FILING DATE:
; APPLICATION NUMBER: AU PM3223
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; REFERENCE/DOCKET NUMBER: 530547/PAS/MKR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215)563-4100
; TELEFAX: (215)563-4044
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-223-043-14

Query Match 91.3%; Score 116; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.1e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAGHGVTSAP 21

```





Query Match 85.8%; Score 109; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGV TSA 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 PDTRPAPGSTAPPAHGV TSA 20

## RESULT 15

US-08-288-059-32  
; Sequence 32, Application US/08288059  
; Patent No. 5827666  
; GENERAL INFORMATION:  
; APPLICANT: FINN, OLIVERA J.  
; APPLICANT: FONTENOT, J. D.  
; APPLICANT: MONTELARO, RONALD C.  
; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN  
; TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/288,059  
; FILING DATE: 08-AUG-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CHAPIN, MARLANA K.  
; REGISTRATION NUMBER: 35,843  
; REFERENCE/DOCKET NUMBER: 61137/205204  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-861-3711  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-288-059-32

Query Match 85.8%; Score 109; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGV TSA 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 PDTRPAPGSTAPPAHGV TSA 20

## RESULT 16

US-08-902-516-20  
; Sequence 20, Application US/08902516  
; Patent No. 5891432  
; GENERAL INFORMATION:  
; APPLICANT: Soo Hoo, William  
; TITLE OF INVENTION: MEMBRANE-BOUND CYTOKINE COMPOSITIONS  
; TITLE OF INVENTION: COMPRISING GM-CSF AND METHODS OF MODULATING AN IMMUNE  
; TITLE OF INVENTION: RESPONSE USING SAME  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CAMPBELL & FLORES, LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/902,516  
; FILING DATE: 29-JUL-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-IM 2442  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619)535-9001  
; TELEFAX: (619)535-8949  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-902-516-20

Query Match 85.8%; Score 109; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGV TSA 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 PDTRPAPGSTAPPAHGV TSA 20

## RESULT 17

US-09-339-944-1  
; Sequence 1, Application US/09339944  
; Patent No. 6114129  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, Babita  
; APPLICANT: LONGENECKER, B. Michael  
; TITLE OF INVENTION: METHODS OF DETECTING T-CELL ACTIVATION AND TREATING  
; TITLE OF INVENTION: DISORDERS ASSOCIATED WITH T-CELL DYSFUNCTION  
; FILE REFERENCE: 042881/0129  
; CURRENT APPLICATION NUMBER: US/09/339,944  
; CURRENT FILING DATE: 1999-06-25  
; EARLIER APPLICATION NUMBER: 60/090,916  
; EARLIER FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-339-944-1

Query Match 85.8%; Score 109; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGV TSA 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 PDTRPAPGSTAPPAHGV TSA 20

## RESULT 18

US-08-737-896-3  
; Sequence 3, Application US/08737896

```
; Patent No. 6168804
; GENERAL INFORMATION:
; APPLICANT: Samuel, John
; APPLICANT: Kwon, Glen S.
; TITLE OF INVENTION: METHOD FOR ELICITING TH1-SPECIFIC
; TITLE OF INVENTION: IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,896
; FILING DATE: 24-SEP-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,499
; FILING DATE: 07-JUN-1996
; APPLICATION NUMBER: PCT/US96/09551
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07254/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SP1-023 MUC1
; US-08-737-896-3

Query Match 85.8%; Score 109; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTS 20
Db 1 PDTRPAPGSTAPPAHGVTS 20

RESULT 19
US-08-134-198E-34
; Sequence 34, Application US/08134198E
; Patent No. 6190885
; GENERAL INFORMATION:
; APPLICANT: CANCER RESEARCH FUND
; APPLICANT: OF CONTRA COSTA
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: FUSION PROTEIN CONTAINING HMEG
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower Street, Suite 1900
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/134,198E
; FILING DATE: October 8, 1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38208 (CRFC-003C)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 622-7700
; TELEFAX: (213) 489-4210
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-134-198E-34

Query Match 85.8%; Score 109; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTS 20
Db 1 PDTRPAPGSTAPPAHGVTS 20

RESULT 20
US-09-847-185-20
; Sequence 20, Application US/09847185
; Patent No. 6482407
; GENERAL INFORMATION:
; APPLICANT: Soc Hoo, William
; TITLE OF INVENTION: MEMBRANE-BOUND CYTOKINE COMPOSITIONS
; COMPRISING GM-CSF AND METHODS OF MODULATING AN IMMUNE
; RESPONSE USING SAME
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL & FLORES, LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/847,185
; FILING DATE: 01-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/201,931
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-IM 2442
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)535-9001
; TELEFAX: (619)535-8949
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-847-185-20

Query Match      85.8%; Score 109; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTS 20
Db 1 PDTRPAPGSTAPPAGHGVTS 20

RESULT 21
US-09-646-028-40
; Sequence 40, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-40

Query Match      85.8%; Score 109; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTS 20
Db 1 PDTRPAPGSTAPPAGHGVTS 20

RESULT 22
US-09-497-232-9
; Sequence 9, Application US/09497232
; Patent No. 6600012
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; KRANTZ, Mark J.
; REDDISH, Mark A.
; LONGENECKER, B. Michael
; TITLE OF INVENTION: METHOD FOR GENERATING ACTIVATED T-CELLS
; AND ANTIGEN-PULSED ANTIGEN-PRESENTING CELLS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/497,232
```

```
;
; FILING DATE: 03-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,410
; FILING DATE: 08-MAY-1998
; APPLICATION NUMBER: US 60/045,949
; FILING DATE: 08-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 042881/0114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-497-232-9

Query Match      85.8%; Score 109; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTS 20
Db 1 PDTRPAPGSTAPPAGHGVTS 20

RESULT 23
US-09-651-265-1
; Sequence 1, Application US/09651265
; Patent No. 6602660
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, B. MICHAEL
; TITLE OF INVENTION: METHODS OF DETECTING T-CELL ACTIVATION
; FILE REFERENCE: 042881/0151
; CURRENT APPLICATION NUMBER: US/09/651,265
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 09/339,344
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/090,916
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-651-265-1

Query Match      85.8%; Score 109; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTS 20
Db 1 PDTRPAPGSTAPPAGHGVTS 20

RESULT 24
US-09-000-003A-3
; Sequence 3, Application US/09000003A
; Patent No. 6652850
; GENERAL INFORMATION:
```

```
; APPLICANT: Philip, Ramila
; Lebkowski, Jane S.
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRAL LIPOSOMES AND
; THEIR USE IN TRANSFECTING DENDRITIC CELLS TO STIMULATE
; SPECIFIC IMMUNITY
;
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alexis Barron, Esq.
; STREET: Suite 2600 Aramark Tower, 1101 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: United States of America
; ZIP: 19107
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,003A
; FILING DATE: 15-Jun-1998
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/12012
; FILING DATE: 19-JUL-1996
; APPLICATION NUMBER: US 60/001,312
; FILING DATE: 21-JUL-1995
; APPLICATION NUMBER: US 60/007,184
; FILING DATE: 01-NOV-1995
; APPLICATION NUMBER: US 08/566,286
; FILING DATE: 01-DEC-1995
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Barron, Alexis
; REGISTRATION NUMBER: 22,702
; REFERENCE/DOCKET NUMBER: 20,846-K USA
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 923-4466
; TELEFAX: (215) 923-2189
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-000-003A-3

Query Match      85.8%; Score 109; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 PDTRPAPGSTAPPAGHGTSA 20
      |||||
Db      1 PDTRPAPGSTAPPAGHGTSA 20

RESULT 25
US-09-601-729-140
; Sequence 140, Application US/09601729
; Patent No. 6683052
;
; GENERAL INFORMATION:
; APPLICANT: THIAM, KADER
; APPLICANT: AURIAULT, CLAUDE
; APPLICANT: GRAS-MASSE, HELENE
; APPLICANT: LOING, ESTELLE
; APPLICANT: VERWAERDE, CLAUDIE
; APPLICANT: GUILLET, JEAN GERARD
;
; TITLE OF INVENTION: LIPOPEPTIDES CONTAINING AN INTERFERON FRAGMENT AND USES
; TITLE OF INVENTION: THEREOF IN PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: USB-97-AU-IN
; CURRENT APPLICATION NUMBER: US/09/601,729
; CURRENT FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: PCT/FR99/00259
```

```
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: 98 01439
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 281
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 140
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-601-729-140

Query Match      85.8%; Score 109; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 PDTRPAPGSTAPPAGHGTSA 20
      |||||
Db      1 PDTRPAPGSTAPPAGHGTSA 20

RESULT 26
PCT-US96-09951-3
; Sequence 3, Application PC/TUS9609951
; GENERAL INFORMATION:
; APPLICANT: The Governors of the University of Alberta
; TITLE OF INVENTION: A METHOD FOR ELICITING A Th1-SPECIFIC
; TITLE OF INVENTION: IMMUNE RESPONSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09951
; FILING DATE: 06-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Learn, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07254/037W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: SP1-023 MUC1
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..20
PCT-US96-09951-3

Query Match      85.8%; Score 109; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 PDTRPAPGSTAPPAGHGTSA 20
```

Db 1 PDTRPAPGSTAPPAHGVTSA 20  
|||||  
RESULT 27  
US-08-288-059-28  
; Sequence 28, Application US/08288059  
; Patent No. 5827666  
; GENERAL INFORMATION:  
; APPLICANT: FINN, OLIVERA J.  
; APPLICANT: FONTELOT, J. D.  
; APPLICANT: MONTELLARO, RONALD C.  
; TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN  
; AND MUCIN-LIKE PEPTIDES, AND USES THEREOF  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/288,059  
; FILING DATE: 08-AUG-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CHAPIN, MARLANA K.  
; REGISTRATION NUMBER: 35,843  
; REFERENCE/DOCKET NUMBER: 61137/205204  
; TELEPHONE: 202-861-3711  
; TELEFAX: 202-822-0944  
; TELEX: 5714627 CUSH  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-288-059-28  
Query Match 85.8%; Score 109; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PDTRPAPGSTAPPAHGVTSA 20  
Db 6 PDTRPAPGSTAPPAHGVTSA 25  
|||||  
RESULT 28  
US-08-737-896-6  
; Sequence 6, Application US/08737896  
; Patent No. 6168804  
; GENERAL INFORMATION:  
; APPLICANT: Samuel, John  
; APPLICANT: Kwon, Glen S.  
; TITLE OF INVENTION: METHOD FOR ELICITING TH1-SPECIFIC  
; IMMUNE RESPONSE  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA

ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/737,896  
; FILING DATE: 24-SEP-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/480,499  
; FILING DATE: 07-JUN-1996  
; APPLICATION NUMBER: PCT/US96/09551  
; FILING DATE: 07-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07254/037001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; CLONE: SP1-070 MUC1  
US-08-737-896-6  
Query Match 85.8%; Score 109; DB 3; Length 30;  
Best Local Similarity 100.0%; Pred. No. 4.3e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 RPAPGSTAPPAHGVTSA 23  
Db 1 RPAPGSTAPPAHGVTSA 20  
|||||  
RESULT 29  
PCT-US96-09951-6  
; Sequence 6, Application PC/TUS9609951  
; GENERAL INFORMATION:  
; APPLICANT: The Governors of the University of Alberta  
; TITLE OF INVENTION: A METHOD FOR ELICITING A TH1-SPECIFIC  
; IMMUNE RESPONSE  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/09951  
; FILING DATE: 06-JUN-1996  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Learn, June M.  
; REGISTRATION NUMBER: 31,238  
; REFERENCE/DOCKET NUMBER: 07254/037001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:



```
/ LENGTH: 30 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ IMMEDIATE SOURCE:
/ CLONE: SPL-070 MUC1
/ FEATURE:
/ NAME/KEY: Peptide
/ LOCATION: 1..30
/ PCT-US96-09951-6

Query Match      85.8%; Score 109; DB 5; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 RPAPGSTAPPAHGVTSAPDT 23
Db 1 RPAPGSTAPPAHGVTSAPDT 20
|||||

RESULT 30
US-08-479-537A-5
/ Sequence 5, Application US/08479537A
/ Patent No. 5861381
/ GENERAL INFORMATION:
/ APPLICANT: CHAMBON, Pierre
/ APPLICANT: KIENY, Marie-Paule
/ APPLICANT: LATHE, Richard
/ APPLICANT: HAREUVENI, Mara
/ TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
/ TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
/ NUMBER OF SEQUENCES: 5
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
/ STREET: P.O. Box 1404
/ CITY: Alexandria
/ STATE: Virginia
/ COUNTRY: United States
/ ZIP: 22313-1404
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/479,537A
/ FILING DATE: 07-JUN-1995
/ CLASSIFICATION: 514
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: FR 90/13101
/ FILING DATE: 23-OCT-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: WO PCT/FR91/00835
/ FILING DATE: 23-OCT-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/039,320
/ FILING DATE: 04-APR-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/403,576
/ FILING DATE: 14-MAR-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Teskin, Robin L.
/ REGISTRATION NUMBER: 35,030
/ REFERENCE/DOCKET NUMBER: 017753-025
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 836-6620
/ TELEFAX: (703) 836-2021
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1867 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
```

```
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ FEATURE:
/ NAME/KEY: Peptide
/ LOCATION: 128..1727
/ OTHER INFORMATION: /note= "The amino acids spanning
/ OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat :
/ OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
/ OTHER INFORMATION: repeats varies from 1 to 40."
/ FEATURE:
/ NAME/KEY: Peptide
/ LOCATION: 134
/ OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
/ OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
/ OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
/ FEATURE:
/ NAME/KEY: Peptide
/ LOCATION: 144
/ OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
/ OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
/ OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
/ FEATURE:
/ NAME/KEY: Peptide
/ LOCATION: 147
/ OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
/ OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
/ OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
/ FEATURE:
/ NAME/KEY: Peptide
/ LOCATION: 1..21
/ OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
/ OTHER INFORMATION: 21 amino acid precursor sequence."
/ US-08-479-537A-5

Query Match      82.7%; Score 105; DB 2; Length 1867;
Best Local Similarity 86.4%; Pred. No. 9.8e-05;
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPD 22
Db 122 PDNKPAPGSTAPXAHGVTSAPD 143
|||||

RESULT 31
US-09-083-116-5
/ Sequence 5, Application US/09083116
/ Patent No. 6203795
/ GENERAL INFORMATION:
/ APPLICANT: CHAMBON, Pierre
/ APPLICANT: KIENY, Marie-Paule
/ APPLICANT: LATHE, Richard
/ APPLICANT: HAREUVENI, Mara
/ TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
/ TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
/ NUMBER OF SEQUENCES: 5
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
/ STREET: P.O. Box 1404
/ CITY: Alexandria
/ STATE: Virginia
/ COUNTRY: United States
/ ZIP: 22313-1404
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/083,116
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/479,537
```

```

; FILING DATE:
; PRIOR APPLICATION DATA: MO PCT/FR91/00835
; APPLICATION NUMBER: 23-OCT-1991
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 128..1727
; OTHER INFORMATION: /note= "The amino acids spanning
; OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat i
; OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
; OTHER INFORMATION: repeats varies from 1 to 40."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 134
; OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 144
; OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 147
; OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..21
; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
; OTHER INFORMATION: 21 amino acid precursor sequence."
; US-09-083-116-5
;
; Query Match 82.7%; Score 105; DB 3; Length 1867;
; Best Local Similarity 86.4%; Pred. No. 9.8e-05;
; Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
;
; Qy 1 PDTRPAPGSTAPPAGHGVTSAPD 22
; ||:|||||
; Db 122 PDNKPAPGSTAPPAGHGVTSAPD 143
;
; RESULT 32
; US-09-134-916A-5
; Sequence 5, Application US/09134916A
; Patent No. 6328956
; GENERAL INFORMATION:
; APPLICANT: CHAMBERN, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard

```

```

; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,916A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 128..1727
; OTHER INFORMATION: /note= "The amino acids spanning
; OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
; OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
; OTHER INFORMATION: repeats varies from 1 to 40."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 134
; OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 144
; OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, AC
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 147
; OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CC
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

```

```
;
;
; NAME/KEY: Peptide
; LOCATION: 1..21
; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
; OTHER INFORMATION: 21 amino acid precursor sequence."
;
; US-09-134-916A-5
;
; Query Match      82.7%; Score 105; DB 4; Length 1867;
; Best Local Similarity 86.4%; Pred. No. 9.8e-05;
; Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
;
; QY      1 PDTRPAPGSTAPPAHGVTSAPD 22
;         ||:||||| ||:|||||
; Db      122 PDNKPAPGSTAPXAHGVTSAPD 143
;
; RESULT 33
; US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 128..1899
```

```
;
;
; OTHER INFORMATION: /note= "The amino acids spanning
; OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat :
; OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
; OTHER INFORMATION: repeats varies from 1 to 40."
;
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 134
; OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
; OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
; OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 144
; OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACY
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
;
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 147
; OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
;
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..21
; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
; OTHER INFORMATION: 21 amino acid precursor sequence."
;
; US-08-479-537A-2
;
; Query Match      82.7%; Score 105; DB 2; Length 2035;
; Best Local Similarity 86.4%; Pred. No. 0.00011;
; Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
;
; QY      1 PDTRPAPGSTAPPAHGVTSAPD 22
;         ||:||||| ||:|||||
; Db      122 PDNKPAPGSTAPXAHGVTSAPD 143
;
; RESULT 34
; US-09-083-116-2
; Sequence 2, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,537
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
```



```
US-09-134-916A-2
Query Match      82.7%; Score 105; DB 4; Length 2035;
Best Local Similarity 86.4%; Pred. No. 0.00011;
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGVTSAPD 22
   ||:||||||| |||||||||
Db 122 PDNKPAPGSTAPXAHGVTSAPD 143

RESULT 36
US-09-646-028-1
; Sequence 1, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-1

Query Match      81.9%; Score 104; DB 4; Length 134;
Best Local Similarity 95.0%; Pred. No. 8.2e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGVTS 20
   ||||||||| |||||||||
Db 112 PDTRPAPGSTPPPAHGVTS 131

RESULT 37
US-09-646-028-2
; Sequence 2, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-2

Query Match      81.9%; Score 104; DB 4; Length 137;
Best Local Similarity 95.0%; Pred. No. 8.4e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGVTS 20
```

```
Db 115 PDTRPAPGSTPPPAHGVTS 134
   ||||||||| |||||||||

RESULT 38
US-09-646-028-3
; Sequence 3, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-3

Query Match      81.9%; Score 104; DB 4; Length 138;
Best Local Similarity 95.0%; Pred. No. 8.5e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGVTS 20
   ||||||||| |||||||||
Db 116 PDTRPAPGSTPPPAHGVTS 135

RESULT 39
US-09-646-028-4
; Sequence 4, Application US/09646028
; Patent No. 6562347
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/09/646,028
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-09-646-028-4

Query Match      81.9%; Score 104; DB 4; Length 156;
Best Local Similarity 95.0%; Pred. No. 9.7e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAHGVTS 20
   ||||||||| |||||||||
Db 134 PDTRPAPGSTPPPAHGVTS 153

RESULT 40
US-08-833-807-1
; Sequence 1, Application US/08833807
```



Patent No. 5989552  
GENERAL INFORMATION:  
APPLICANT: McKenzie, Ian F.C.  
APPLICANT: Apostolopoulos, Vasso  
APPLICANT: Pietersz, Geoff A.  
TITLE OF INVENTION: ANTIGENIC CARBOHYDRATE COMPOUNDS AND  
TITLE OF INVENTION: THEIR USE IN IMMUNOTHERAPY  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dann Dorfman Herrell and Skillman  
STREET: Suite 720, 1601 Market Street  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: United States of America  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/833,807  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/340,711  
FILING DATE: 16-NOV-1994  
APPLICATION NUMBER: AU PM3223  
FILING DATE: 24-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hagan, Patrick J.  
REGISTRATION NUMBER: 27,643  
REFERENCE/DOCKET NUMBER: 530547/PAS/MKR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215)563-4100  
TELEFAX: (215)563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-833-807-1

Query Match 79.5%; Score 101; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.6e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDTRPAPGSTAPPAGHVT 18  
Db 3 PDTRPAPGSTAPPAGHVT 20

Search completed: May 6, 2004, 16:32:38  
Job time : 20.4615 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:31:15 ; Search time 43.641 Seconds  
(without alignments)  
146.285 Million cell updates/sec

Title: US-10-070-566-7

Perfect score: 127

Sequence: 1 PDTRPAPGSTAPPAHGVTAPDT 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	127	100.0	40	10	US-09-996-069-9	Sequence 9, Appl
2	127	100.0	43	14	US-10-106-876-14	Sequence 14, Appl
3	127	100.0	46	10	US-09-996-069-5	Sequence 5, Appl
4	127	100.0	46	14	US-10-106-876-1	Sequence 1, Appl
5	127	100.0	100	9	US-09-965-131-6	Sequence 6, Appl
6	127	100.0	508	14	US-10-057-136-20	Sequence 20, Appl
7	127	100.0	1255	10	US-09-996-069-10	Sequence 10, Appl
8	127	100.0	1255	14	US-10-171-311-158	Sequence 158, App
9	127	100.0	1255	14	US-10-177-293-311	Sequence 311, App
10	122	96.1	475	12	US-10-247-703-22	Sequence 22, Appl
11	122	96.1	475	15	US-10-417-312-1	Sequence 1, Appl
12	122	96.1	515	12	US-10-612-090-19	Sequence 19, Appl
13	122	96.1	515	12	US-10-247-703-20	Sequence 20, Appl
14	122	96.1	515	14	US-10-097-340-212	Sequence 212, App
15	122	96.1	515	14	US-10-171-311-156	Sequence 156, App

16	119	93.7	156	12	US-10-247-703-37	Sequence 37, Appl
17	113	89.0	172	14	US-10-335-394-49	Sequence 49, Appl
18	113	89.0	177	14	US-10-335-394-54	Sequence 54, Appl
19	113	89.0	307	12	US-10-247-703-26	Sequence 26, Appl
20	109	85.8	20	9	US-09-847-185-20	Sequence 20, Appl
21	109	85.8	20	9	US-09-984-183-11	Sequence 11, Appl
22	109	85.8	20	9	US-09-984-333-1	Sequence 1, Appl
23	109	85.8	20	12	US-10-612-090-3	Sequence 3, Appl
24	109	85.8	20	14	US-10-224-286-20	Sequence 20, Appl
25	109	85.8	20	14	US-10-335-394-40	Sequence 40, Appl
26	109	85.8	20	15	US-10-406-317-31	Sequence 31, Appl
27	109	85.8	21	14	US-10-062-710-196	Sequence 196, App
28	109	85.8	21	14	US-10-062-710-207	Sequence 207, App
29	109	85.8	35	9	US-09-984-183-12	Sequence 12, Appl
30	109	85.8	35	9	US-09-984-183-13	Sequence 13, Appl
31	104	81.9	134	14	US-10-335-394-1	Sequence 1, Appl
32	104	81.9	137	14	US-10-335-394-2	Sequence 2, Appl
33	104	81.9	138	14	US-10-335-394-3	Sequence 3, Appl
34	104	81.9	156	14	US-10-335-394-4	Sequence 4, Appl
35	92	72.4	20	9	US-09-822-698A-8	Sequence 8, Appl
36	92	72.4	125	12	US-10-296-734-823	Sequence 823, App
37	88	69.3	24	12	US-10-296-734-1166	Sequence 1166, Ap
38	88	69.3	5546	12	US-10-296-734-1210	Sequence 1210, Ap
39	86	67.7	20	9	US-09-984-183-16	Sequence 16, Appl
40	86	67.7	20	9	US-09-984-333-6	Sequence 6, Appl
41	86	67.7	20	12	US-10-296-317-45	Sequence 45, Appl
42	86	67.7	20	13	US-10-054-488-1	Sequence 1, Appl
43	86	67.7	20	14	US-10-057-136-1	Sequence 1, Appl
44	86	67.7	20	14	US-10-057-136-16	Sequence 16, Appl
45	86	67.7	36	12	US-10-296-317-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1

US-09-996-069-9  
; Sequence 9, Application US/09996069  
; Publication No. US20030036199A1  
; GENERAL INFORMATION:  
; APPLICANT: Bamdad, Cynthia  
; APPLICANT: Bamdad, R. Shoshana  
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIB.  
; TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER  
; FILE REFERENCE: M01015/70071  
; CURRENT APPLICATION NUMBER: US/09/996,069  
; CURRENT FILING DATE: 2001-11-27  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 9  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-996-069-9

Query Match 100.0%; Score 127; DB 10; Length 40;  
Best local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTAPDT 23

Db 1 PDTRPAPGSTAPPAHGVTAPDT 23

RESULT 2

US-10-106-876-14  
; Sequence 14, Application US/10106876  
; Publication No. US20030157160A1  
; GENERAL INFORMATION:  
; APPLICANT: BUDZYNSKI, WLADYSLAW A.  
; APPLICANT: KOGANTY, R. RAO  
; APPLICANT: KRANTZ, MARK J.  
; APPLICANT: LONGENECKER, B. MICHAEL

```
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE
; TITLE OF INVENTION: RESPONSES
; FILE REFERENCE: 042881-0176
; CURRENT APPLICATION NUMBER: US/10/106,876
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/278,698
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-106-876-14
```

```
Query Match      100.0%; Score 127; DB 14; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 PDTRPAPGSTAPPARGVTSAPDT 23
        |||||
Db      4 PDTRPAPGSTAPPARGVTSAPDT 26
```

## RESULT 3

```
US-09-996-069-5
; Sequence 5, Application US/09996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBITORS, AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-5
```

```
Query Match      100.0%; Score 127; DB 10; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 PDTRPAPGSTAPPARGVTSAPDT 23
        |||||
Db      1 PDTRPAPGSTAPPARGVTSAPDT 23
```

## RESULT 4

```
US-10-106-876-1
; Sequence 1, Application US/10106876
; Publication No. US20030157160A1
; GENERAL INFORMATION:
; APPLICANT: BUDZYNSKI, WLADYSLAW A.
; APPLICANT: KOGANTY, R. RAO
; APPLICANT: KRANTZ, MARK J.
; APPLICANT: LONGENECKER, B. MICHAEL
; TITLE OF INVENTION: VACCINE FOR MODULATING BETWEEN T1 AND T2 IMMUNE
; TITLE OF INVENTION: RESPONSES
; FILE REFERENCE: 042881-0176
; CURRENT APPLICATION NUMBER: US/10/106,876
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/278,698
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 20
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-106-876-1
```

```
Query Match      100.0%; Score 127; DB 14; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 PDTRFAPGSTAPPARGVTSAPDT 23
        |||||
Db      7 PDTRFAPGSTAPPARGVTSAPDT 29
```

## RESULT 5

```
US-09-965-131-6
; Sequence 6, Application US/09965131
; Patent No. US20020160502A1
; GENERAL INFORMATION:
; APPLICANT: Chung, Maureen A.
; APPLICANT: Sharma, Surendra
; APPLICANT: Chag, Helena R.
; APPLICANT: O'Donnell, Mark A.
; TITLE OF INVENTION: RECOMBINANT BCG VACCINES FOR THE PREVENTION AND TREATMENT OF CANCER
; FILE REFERENCE: WII-014CP
; CURRENT APPLICATION NUMBER: US/09/965,131
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,455
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-131-6
```

```
Query Match      100.0%; Score 127; DB 9; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 PDTRFAPGSTAPPARGVTSAPDT 23
        |||||
Db      6 PDTRFAPGSTAPPARGVTSAPDT 28
```

## RESULT 6

```
US-10-057-136-20
; Sequence 20, Application US/10057136
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHLOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFF, DONALD
; APPLICANT: PANICALI, DENNIS
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 09/366,670
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
```

```

; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20

Query Match      100.0%; Score 127; DB 14; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PDTRPAPGSTAPPAHGVTSAPDT 23
      |||||
Db      89 PDTRPAPGSTAPPAHGVTSAPDT 111

RESULT 7
US-09-996-069-10
; Sequence 10, Application US/09996069
; Publication No. US20030036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBIT
; TITLE OF INVENTION: AND COMPOSITIONS AND METHODS FOR TREATMENT OF CANCER
; FILE REFERENCE: M01015/70071
; CURRENT APPLICATION NUMBER: US/09/996,069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-069-10

Query Match      100.0%; Score 127; DB 10; Length 1255;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PDTRPAPGSTAPPAHGVTSAPDT 23
      |||||
Db      142 PDTRPAPGSTAPPAHGVTSAPDT 164

RESULT 8
US-10-171-311-158
; Sequence 158, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoers, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 158
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-158

Query Match      100.0%; Score 127; DB 14; Length 1255;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PDTRPAPGSTAPPAHGVTSAPDT 23
      |||||
Db      142 PDTRPAPGSTAPPAHGVTSAPDT 164

RESULT 9
US-10-177-293-311
; Sequence 311, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: East Jr., Robert C.
; APPLICANT: Hortchbagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-311

Query Match      100.0%; Score 127; DB 14; Length 1255;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PDTRPAPGSTAPPAHGVTSAPDT 23
      |||||
Db      142 PDTRPAPGSTAPPAHGVTSAPDT 164

RESULT 10

```

```

US-10-247-703-22
; Sequence 22, Application US/10247703
; Publication No. US20030063597A1
; GENERAL INFORMATION:
; APPLICANT: Branigan, Patrick
; APPLICANT: Goletz, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: Scallan, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS
; TITLE OF INVENTION: CYTOKINE ADJUVANT ENCODING NUCLEIC ACID
; FILE REFERENCE: CEN310
; CURRENT APPLICATION NUMBER: US/10/247,703
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/328,371
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-247-703-22

Query Match          96.1%; Score 122; DB 12; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPD 22
Db 142 PDTRPAPGSTAPPAHGVTSAPD 163

RESULT 11
US-10-417-312-1
; Sequence 1, Application US/10417312
; Publication No. US20030235868A1
; GENERAL INFORMATION:
; APPLICANT: Dyax Corp
; TITLE OF INVENTION: Antibodies Specific for Mucin Polypeptide
; FILE REFERENCE: 2403/2002
; CURRENT APPLICATION NUMBER: US/10/417,312
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/374,432
; PRIOR FILING DATE: 2002-04-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-417-312-1

Query Match          96.1%; Score 122; DB 15; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPD 22
Db 142 PDTRPAPGSTAPPAHGVTSAPD 163

RESULT 12
US-10-612-090-19
; Sequence 19, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:
; APPLICANT: Immunogen, Inc.
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03

```

```

; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Exemplary Mucl protein
US-10-612-090-19

Query Match          96.1%; Score 122; DB 12; Length 515;
Best Local Similarity 95.7%; Pred. No. 5.1e-06;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPDT 23
Db 142 PDTRPAPGSTAPPAHGVTSAPDT 164

RESULT 13
US-10-247-703-20
; Sequence 20, Application US/10247703
; Publication No. US20030063597A1
; GENERAL INFORMATION:
; APPLICANT: Branigan, Patrick
; APPLICANT: Goletz, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: Scallan, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS
; TITLE OF INVENTION: CYTOKINE ADJUVANT ENCODING NUCLEIC ACID
; FILE REFERENCE: CEN310
; CURRENT APPLICATION NUMBER: US/10/247,703
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/328,371
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-247-703-20

Query Match          96.1%; Score 122; DB 12; Length 515;
Best Local Similarity 95.7%; Pred. No. 5.1e-06;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPDT 23
Db 142 PDTRPAPGSTAPPAHGVTSAPDT 164

RESULT 14
US-10-097-340-212
; Sequence 212, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.

```



```
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 212
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-212
```

Query Match 96.1%; Score 122; DB 14; Length 515;  
Best Local Similarity 95.7%; Pred. No. 5.1e-06;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23
Db 142 PDTRPPPGSTAPPAGHGVTSAPDT 164
```

```
RESULT 15
US-10-171-311-156
; Sequence 156, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-156
```

```
Query Match 96.1%; Score 122; DB 14; Length 515;
Best Local Similarity 95.7%; Pred. No. 5.1e-06;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23
Db 142 PDTRPPPGSTAPPAGHGVTSAPDT 164

RESULT 16
US-10-247-703-37
; Sequence 37, Application US/102477703
; Publication No. US20030063597A1
; GENERAL INFORMATION:
; APPLICANT: Branigan, Patrick
; APPLICANT: Goletz, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: Scallion, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS
; TITLE OF INVENTION: CYTOKINE ADJUVANT ENCODING NUCLEIC ACID
; FILE REFERENCE: CEN310
; CURRENT APPLICATION NUMBER: US/10/247,703
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/328,371
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-247-703-37
```

Query Match 93.7%; Score 119; DB 12; Length 156;  
Best Local Similarity 91.3%; Pred. No. 3.5e-06;  
Matches 21; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23
Db 131 PDNKPAPGSTAPPAGHGVTSAPDT 153
```

```
RESULT 17
US-10-335-394-49
; Sequence 49, Application US/10335394
; Publication No. US20030138452A1
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/10/335,394
; CURRENT FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US/09/646,028
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-10-335-394-49
```

Query Match 89.0%; Score 113; DB 14; Length 172;  
Best Local Similarity 74.2%; Pred. No. 1.9e-05;  
Matches 23; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 PDTRPAPGSTAPPAH-----GVTSAPDT 23  
Db 104 PDTRPAPGSTAPPAHGVTSAAALDGVTSAPDT 134

## RESULT 18

US-10-335-394-54  
; Sequence 54, Application US/10335394  
; Publication No. US20030138452A1  
; GENERAL INFORMATION:  
; APPLICANT: Kwak, Larry  
; APPLICANT: Biragyn, Arya  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF  
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES  
; FILE REFERENCE: 14C14.0316/P  
; CURRENT APPLICATION NUMBER: US/10/335,394  
; CURRENT FILING DATE: 2002-12-31  
; PRIOR APPLICATION NUMBER: US/09/646,028  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 60/077,745  
; PRIOR FILING DATE: 1998-03-12  
; NUMBER OF SEQ ID NOS: 57  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 54  
; LENGTH: 177  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct  
US-10-335-394-54

Query Match 89.0%; Score 113; DB 14; Length 177;  
Best Local Similarity 74.2%; Pred. No. 1.9e-05;  
Matches 23; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 PDTRPAPGSTAPPAH-----GVTSAPDT 23  
Db 109 PDTRPAPGSTAPPAHGVTSAAALDGVTSAPDT 139

## RESULT 19

US-10-247-703-26  
; Sequence 26, Application US/10247703  
; Publication No. US20030063597A1  
; GENERAL INFORMATION:  
; APPLICANT: Branigan, Patrick  
; APPLICANT: Goletz, Theresa J  
; APPLICANT: Knight, David M  
; APPLICANT: McCarthy, Stephen G  
; APPLICANT: Scallion, Bernard J  
; APPLICANT: Snyder, Linda A  
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS  
; TITLE OF INVENTION: CYTOKINE ADJUVANT ENCODING NUCLEIC ACID  
; FILE REFERENCE: CEN310  
; CURRENT APPLICATION NUMBER: US/10/247,703  
; CURRENT FILING DATE: 2002-09-20  
; PRIOR APPLICATION NUMBER: 60/328,371  
; PRIOR FILING DATE: 2001-10-10  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 26  
; LENGTH: 307  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-247-703-26

Query Match 89.0%; Score 113; DB 12; Length 307;  
Best Local Similarity 90.9%; Pred. No. 3.3e-05;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPD 22  
Db 104 PDTRPAPGSTAPPAHGVTSAAALDGVTSAPDT 134

Db 142 PETREPPGSTAPPAHGVTSAPD 163

## RESULT 20

US-09-847-185-20  
; Sequence 20, Application US/09847185  
; Patent No. US20020076392A1  
; GENERAL INFORMATION:  
; APPLICANT: Soo Hoo, William  
; TITLE OF INVENTION: MEMBRANE-BOUND CYTOKINE COMPOSITIONS  
; TITLE OF INVENTION: COMPRISING GM-CSF AND METHODS OF MODULATING AN IMMUNE  
; RESPONSE USING SAME  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CAMPBELL & FLORES, LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/847,185  
; FILING DATE: 01-May-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/201,931  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-IM 2442  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619)535-9001  
; TELEFAX: (619)535-8949  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:  
US-09-847-185-20

Query Match 85.8%; Score 109; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPD 20  
Db 1 PDTRPAPGSTAPPAHGVTSAPD 20

## RESULT 21

US-09-984-183-11  
; Sequence 11, Application US/09984183  
; Patent No. US20020142983A1  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, BABITA  
; APPLICANT: LONGENECKER, MICHAEL B.  
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE  
; TITLE OF INVENTION: DISORDERS  
; FILE REFERENCE: 042881/0130  
; CURRENT APPLICATION NUMBER: US/09/984,183  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 09/457,354  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/111,973  
; PRIOR FILING DATE: 1998-12-11

; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-183-11

Query Match 85.8%; Score 109; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGV TSA 20  
|||||  
Db 1 PDTRPAPGSTAPPAHGV TSA 20

## RESULT 22

US-09-984-333-1  
; Sequence 1, Application US/09384333  
; Patent No. US20020159969A1  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, Babita  
; APPLICANT: LONGENECKER, Bryan Michael  
; APPLICANT: REDDISH, Mark Austin  
; TITLE OF INVENTION: SMALL PEPTIDE-BASED THERAPEUTICS FOR REVERSING  
; FILE REFERENCE: 042881/0116  
; CURRENT APPLICATION NUMBER: US/09/984,333  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 09/182,887  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: 60/064,146  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/065,209  
; PRIOR FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-333-1

Query Match 85.8%; Score 109; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGV TSA 20  
|||||  
Db 1 PDTRPAPGSTAPPAHGV TSA 20

## RESULT 23

US-10-612-090-3  
; Sequence 3, Application US/10612090  
; Publication No. US20040057952A1  
; GENERAL INFORMATION:  
; APPLICANT: ImmunoGen, Inc.  
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF  
; FILE REFERENCE: A8340  
; CURRENT APPLICATION NUMBER: US/10/612,090  
; CURRENT FILING DATE: 2003-07-03  
; PRIOR APPLICATION NUMBER: US 60/393,094  
; PRIOR FILING DATE: 2002-07-03  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-612-090-3

Query Match 85.8%; Score 109; DB 12; Length 20;  
Best Local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGV TSA 20  
|||||  
Db 1 PDTRPAPGSTAPPAHGV TSA 20

## RESULT 24

US-10-224-286-20  
; Sequence 20, Application US/10224286  
; Publication No. US20030108517A1  
; GENERAL INFORMATION:  
; APPLICANT: SOO HOO, William  
; TITLE OF INVENTION: MEMBRANE-BOUND CYTOKINE COMPOSITIONS  
; COMPRISING GM-CSF AND METHODS OF MODULATING AN IMMUNE  
; RESPONSE USING SAME  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CAMPBELL & FLORES, LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/224,286  
; FILING DATE: 19-Aug-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/902,516  
; FILING DATE: 29-JUL-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-IM 2442  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619)535-9001  
; TELEFAX: (619)535-8949  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:  
US-10-224-286-20

Query Match 85.8%; Score 109; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGV TSA 20  
|||||  
Db 1 PDTRPAPGSTAPPAHGV TSA 20

## RESULT 25

US-10-335-394-40  
; Sequence 40, Application US/10335394  
; Publication No. US20030138452A1  
; GENERAL INFORMATION:  
; APPLICANT: Kwak, Larry  
; APPLICANT: Biragyn, Arya  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF  
; CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES  
; FILE REFERENCE: 14014.0316/P

; CURRENT APPLICATION NUMBER: US/10/335,394  
; CURRENT FILING DATE: 2002-12-31  
; PRIOR APPLICATION NUMBER: US/09/646,028  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 60/077,745  
; PRIOR FILING DATE: 1998-03-12  
; NUMBER OF SEQ ID NOS: 57  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 40  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct  
US-10-335-394-40

Query Match 85.8%; Score 109; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTS 20  
Db 1 PDTRPAPGSTAPPAHGVTS 20

## RESULT 26

US-10-406-317-31  
; Sequence 31, Application US/10406317  
; Publication No. US20040019195A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlom, Jeffrey;  
; APPLICANT: Hodge, James;  
; APPLICANT: Panicali, Dennis  
; TITLE OF INVENTION: A recombinant vector expressing multiple costimulatory  
; TITLE OF INVENTION: molecules and uses thereof  
; FILE REFERENCE: 38163-0189  
; CURRENT APPLICATION NUMBER: US/10/406,317  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: US/09/856,988  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: PCT/US99/26866  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/111,582  
; PRIOR FILING DATE: 1998-12-09  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 31  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-10-406-317-31

Query Match 85.8%; Score 109; DB 15; Length 20;  
Best Local Similarity 100.0%; Pred. No. 6.5e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTS 20  
Db 1 PDTRPAPGSTAPPAHGVTS 20

## RESULT 27

US-10-062-710-196  
; Sequence 196, Application US/10062710  
; Publication No. US20030049253A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Frank Q.  
; APPLICANT: Chu, Yong-Liang  
; APPLICANT: Qiu, Jian-Tai  
; TITLE OF INVENTION: Polymeric Conjugates for Delivery of

; TITLE OF INVENTION: MHC-Recognized Epitopes  
; TITLE OF INVENTION: Via Peptide Vaccines  
; FILE REFERENCE: 3781-001-27  
; CURRENT APPLICATION NUMBER: US/10/062,710  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/310,498  
; PRIOR FILING DATE: 2001-08-08  
; NUMBER OF SEQ ID NOS: 232  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 196  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HER2/neu breast cancer antigens  
US-10-062-710-196

Query Match 85.8%; Score 109; DB 14; Length 21;  
Best Local Similarity 100.0%; Pred. No. 6.8e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTS 20  
Db 2 PDTRPAPGSTAPPAHGVTS 21

## RESULT 28

US-10-062-710-207  
; Sequence 207, Application US/10062710  
; Publication No. US20030049253A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Frank Q.  
; APPLICANT: Chu, Yong-Liang  
; APPLICANT: Qiu, Jian-Tai  
; TITLE OF INVENTION: Polymeric Conjugates for Delivery of  
; TITLE OF INVENTION: MHC-Recognized Epitopes  
; TITLE OF INVENTION: Via Peptide Vaccines  
; FILE REFERENCE: 3781-001-27  
; CURRENT APPLICATION NUMBER: US/10/062,710  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/310,498  
; PRIOR FILING DATE: 2001-08-08  
; NUMBER OF SEQ ID NOS: 232  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 207  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: T Cell epitopes  
US-10-062-710-207

Query Match 85.8%; Score 109; DB 14; Length 21;  
Best Local Similarity 100.0%; Pred. No. 6.8e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTS 20  
Db 2 PDTRPAPGSTAPPAHGVTS 21

## RESULT 29

US-09-984-183-12  
; Sequence 12, Application US/09984183  
; Patent No. US20020142983A1  
; GENERAL INFORMATION:  
; APPLICANT: AGRAWAL, BABITA  
; APPLICANT: LONGENECKER, MICHAEL B.  
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE  
; TITLE OF INVENTION: DISORDERS  
; FILE REFERENCE: 042881/0130  
; CURRENT APPLICATION NUMBER: US/09/984,183  
; CURRENT FILING DATE: 2001-10-29

```
; PRIOR APPLICATION NUMBER: 09/457,354
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/111,973
; PRIOR FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RBS
; LOCATION: (21)..(25)
; OTHER INFORMATION: this region may encompass one to five
; OTHER INFORMATION: amino acids selected from cysteine, serine
; OTHER INFORMATION: threonine and methionine
; NAME/KEY: MOD_RES
; LOCATION: (29)..(30)
; OTHER INFORMATION: this region may encompass one or two
; OTHER INFORMATION: amino acids selected from cysteine, serine,
; OTHER INFORMATION: threonine and methionine
; NAME/KEY: MOD_RES
; LOCATION: (31)..(35)
; OTHER INFORMATION: this region may encompass one to five
; OTHER INFORMATION: amino acids selected from cysteine, serine,
; OTHER INFORMATION: threonine and methionine
US-09-984-183-12
Query Match      85.8%; Score 109; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PDTRPAPGSTAPPAGHGVTS A 20
Db      1 PDTRPAPGSTAPPAGHGVTS A 20

RESULT 30
US-09-984-183-13
; Sequence 13, Application US/09984183
; Patent No. US20020142983A1
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, MICHAEL B.
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 042881/0130
; CURRENT APPLICATION NUMBER: US/09/984,183
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/457,354
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/111,973
; PRIOR FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(5)
; OTHER INFORMATION: this region may encompass one to five
; OTHER INFORMATION: amino acids selected from cysteine, serine
; OTHER INFORMATION: threonine and methionine
; NAME/KEY: MOD_RES
; LOCATION: (9)..(10)
; OTHER INFORMATION: this region may encompass one or two
; OTHER INFORMATION: amino acids selected from cysteine, serine,
; OTHER INFORMATION: threonine and methionine
; NAME/KEY: MOD_RES
; LOCATION: (11)..(15)
; OTHER INFORMATION: this region may encompass one to five

US-09-984-183-13
Query Match      85.8%; Score 109; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PDTRPAPGSTAPPAGHGVTS A 20
Db      1 PDTRPAPGSTAPPAGHGVTS A 20

RESULT 30
US-09-984-183-13
; Sequence 13, Application US/09984183
; Patent No. US20020142983A1
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, MICHAEL B.
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 042881/0130
; CURRENT APPLICATION NUMBER: US/09/984,183
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/457,354
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/111,973
; PRIOR FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(5)
; OTHER INFORMATION: this region may encompass one to five
; OTHER INFORMATION: amino acids selected from cysteine, serine
; OTHER INFORMATION: threonine and methionine
; NAME/KEY: MOD_RES
; LOCATION: (9)..(10)
; OTHER INFORMATION: this region may encompass one or two
; OTHER INFORMATION: amino acids selected from cysteine, serine,
; OTHER INFORMATION: threonine and methionine
; NAME/KEY: MOD_RES
; LOCATION: (11)..(15)
; OTHER INFORMATION: this region may encompass one to five

; OTHER INFORMATION: amino acids selected from cysteine, serine,
; OTHER INFORMATION: threonine and methionine
US-09-984-183-13
Query Match      85.8%; Score 109; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PDTRPAPGSTAPPAGHGVTS A 20
Db      16 PDTRPAPGSTAPPAGHGVTS A 35

RESULT 31
US-10-335-394-1
; Sequence 1, Application US/10335394
; Publication No. US20030138452A1
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/10/335,394
; CURRENT FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US/09/646,028
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-10-335-394-1
Query Match      81.9%; Score 104; DB 14; Length 134;
Best Local Similarity 95.0%; Pred. No. 0.00015;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 PDTRPAPGSTAPPAGHGVTS A 20
Db      112 PDTRPAPGSTAPPAGHGVTS A 131

RESULT 32
US-10-335-394-2
; Sequence 2, Application US/10335394
; Publication No. US20030138452A1
; GENERAL INFORMATION:
; APPLICANT: Kwak, Larry
; APPLICANT: Biragyn, Arya
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES
; FILE REFERENCE: 14014.0316/P
; CURRENT APPLICATION NUMBER: US/10/335,394
; CURRENT FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US/09/646,028
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/077,745
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct
US-10-335-394-2
```



Query Match 81.9%; Score 104; DB 14; Length 137;  
Best Local Similarity 95.0%; Pred. No. 0.00016;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTS 20  
Db 115 PDTRPAPGSTAPPAGHGVTS 134

RESULT 33  
US-10-335-394-3  
; Sequence 3, Application US/10335394  
; Publication No. US20030138452A1  
; GENERAL INFORMATION:  
; APPLICANT: Kwak, Larry  
; APPLICANT: Biragyn, Arya  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF  
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES  
; FILE REFERENCE: 14014.0316/P  
; CURRENT APPLICATION NUMBER: US/10/335,394  
; CURRENT FILING DATE: 2002-12-31  
; PRIOR APPLICATION NUMBER: US/09/646,028  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 60/077,745  
; PRIOR FILING DATE: 1998-03-12  
; NUMBER OF SEQ ID NOS: 57  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 138  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct  
US-10-335-394-3

Query Match 81.9%; Score 104; DB 14; Length 138;  
Best Local Similarity 95.0%; Pred. No. 0.00016;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTS 20  
Db 116 PDTRPAPGSTAPPAGHGVTS 135

RESULT 34  
US-10-335-394-4  
; Sequence 4, Application US/10335394  
; Publication No. US20030138452A1  
; GENERAL INFORMATION:  
; APPLICANT: Kwak, Larry  
; APPLICANT: Biragyn, Arya  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS OF  
; TITLE OF INVENTION: CHEMOKINE-TUMOR ANTIGEN FUSION PROTEINS AS CANCER VACCINES  
; FILE REFERENCE: 14014.0316/P  
; CURRENT APPLICATION NUMBER: US/10/335,394  
; CURRENT FILING DATE: 2002-12-31  
; PRIOR APPLICATION NUMBER: US/09/646,028  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 60/077,745  
; PRIOR FILING DATE: 1998-03-12  
; NUMBER OF SEQ ID NOS: 57  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 156  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of artificial sequence:/note=synthetic construct  
US-10-335-394-4

Query Match 81.9%; Score 104; DB 14; Length 156;  
Best Local Similarity 95.0%; Pred. No. 0.00018;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTS 20  
Db 134 PDTRPAPGSTAPPAGHGVTS 153

RESULT 35  
US-09-822-698A-8  
; Sequence 8, Application US/09822698A  
; Patent No. US20020146750A1  
; GENERAL INFORMATION:  
; APPLICANT: Hoogenboom, Hendricus R.J.M.  
; APPLICANT: Henderikx, Maria P.G.  
; TITLE OF INVENTION: MUCIN-1 Specific Binding Members and Methods of Use Thereof  
; FILE REFERENCE: DYX-015.1 US  
; CURRENT APPLICATION NUMBER: US/09/822,698A  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: US 09/538,913  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: Microsoft Word  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: peptide of MUC1 protein  
US-09-822-698A-8

Query Match 72.4%; Score 92; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.00055;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGH 16  
Db 5 PDTRPAPGSTAPPAGH 20

RESULT 36  
US-10-296-734-823  
; Sequence 823, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A  
; TITLE OF INVENTION: Synthetic molecules and uses therefor  
; FILE REFERENCE: Savine  
; CURRENT APPLICATION NUMBER: US/10/296,734  
; CURRENT FILING DATE: 2003-08-04  
; PRIOR APPLICATION NUMBER: AU PQ7761/00  
; PRIOR FILING DATE: 2000-05-26  
; NUMBER OF SEQ ID NOS: 1507  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 823  
; LENGTH: 125  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: MUC1F consensus polypeptide  
US-10-296-734-823

Query Match 72.4%; Score 92; DB 12; Length 125;  
Best Local Similarity 81.8%; Pred. No. 0.0033;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22  
Db 102 PVTRPAGSTTPPAHDVTSAPD 123

RESULT 37  
US-10-296-734-1166

```
; Sequence 1166, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1166
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: MUC1F segment 8
US-10-296-734-1166
```

```
Query Match 69.3%; Score 88; DB 12; Length 24;
Best Local Similarity 85.0%; Pred. No. 0.0019;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 TRPAPGSTAPPAGHGVTSAPD 22
Db 1 TRPALGSTTPPAHDVTSAPD 20
```

# RESULT 38

```
US-10-296-734-1210
; Sequence 1210, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1210
; LENGTH: 5546
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Differentiation svaine
US-10-296-734-1210
```

```
Query Match 69.3%; Score 88; DB 12; Length 5546;
Best Local Similarity 85.0%; Pred. No. 0.37;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 TRPAPGSTAPPAGHGVTSAPD 22
Db 1411 TRPALGSTTPPAHDVTSAPD 1430
```

# RESULT 39

```
US-09-984-183-16
; Sequence 16, Application US/09984183
; Patent No. US20020142983A1
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, BABITA
; APPLICANT: LONGENECKER, MICHAEL B.
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE
; DISORDERS
; FILE REFERENCE: 042881/0130
; CURRENT APPLICATION NUMBER: US/09/984,183
```

```
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/457,354
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/111,973
; PRIOR FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-183-16
```

```
Query Match 67.7%; Score 86; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 PDTRPAPGSTAPPAH 15
Db 6 PDTRPAPGSTAPPAH 20
```

# RESULT 40

```
US-09-984-333-6
; Sequence 6, Application US/09984333
; Patent No. US20020159969A1
; GENERAL INFORMATION:
; APPLICANT: AGRAWAL, Babita
; APPLICANT: LONGENECKER, Bryan Michael
; APPLICANT: REDDISH, Mark Austin
; TITLE OF INVENTION: SMALL PEPTIDE-BASED THERAPEUTICS FOR REVERSING
; TITLE OF INVENTION: CANCER-ASSOCIATED MUC-1 MUCIN-INDUCED IMMUNOSUPPRESSION
; FILE REFERENCE: 042881/0116
; CURRENT APPLICATION NUMBER: US/09/984,333
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 09/182,887
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 60/064,146
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065,209
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-333-6
```

```
Query Match 67.7%; Score 86; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 PDTRPAPGSTAPPAH 15
Db 6 PDTRPAPGSTAPPAH 20
```

```
Search completed: May 6, 2004, 16:40:45
Job time : 44.641 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: May 6, 2004, 16:26:14 ; Search time 15.9231 Seconds  
(without alignments)  
138.943 Million cell updates/sec

Title: US-10-070-566-7  
Perfect score: 127  
Sequence: 1 PDTRAPGSTAPPAHGVTAPDT 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	100.0	1344	1 A35175	mucin 1 precursor,
2	122	96.1	347	2 S10571	mucin 1 precursor,
3	103	81.1	256	2 A60533	tumor-associated a
4	61	48.0	955	2 S52959	male-specific leth
5	61	48.0	1334	2 T50568	probable multi-dom
6	59	46.5	652	2 S71753	repellent protein
7	57	44.9	544	2 A71260	hypothetical prote
8	57	44.9	730	1 I52580	gelatinase B (EC 3
9	56.5	44.5	407	2 T50505	hypothetical prote
10	56	44.1	218	2 S52286	NADH2 dehydrogenas
11	56	44.1	708	2 JC4364	gelatinase B (EC 3
12	56	44.1	708	2 S62907	gelatinase B (EC 3
13	55.5	43.7	569	2 A11347	hypothetical prote
14	55	43.3	707	1 A34458	gelatinase B (EC 3
15	55	43.3	916	2 T20909	hypothetical prote
16	54	42.5	576	2 T36729	probable serine/th
17	54	42.5	730	2 JC1456	gelatinase B (EC 3
18	54	42.5	927	2 T00357	hypothetical prote
19	52	40.9	115	2 F72570	hypothetical prote
20	52	40.9	196	2 T51838	blue copper bindin
21	52	40.9	576	2 A49933	proteinase SM tran
22	52	40.9	586	2 T48672	ABC-type transport
23	52	40.9	630	2 A39344	tumor-associated m
24	52	40.9	631	2 I52257	episialin - mouse
25	52	40.9	912	2 B97566	hypothetical prote
26	52	40.9	1008	2 AH2786	conserved hypothet
27	52	40.9	2706	2 T28155	variant-specific s
28	52	40.9	3345	2 T13423	hypothetical prote
29	51.5	40.6	204	2 A48725	nuclear DNA-bindin

30	51	40.2	243	2 S28444	tonB protein - Pse
31	51	40.2	374	2 T16486	hypothetical prote
32	51	40.2	571	2 C75530	conserved hypothet
33	51	40.2	594	2 T12995	pectinesterase hom
34	51	40.2	698	2 T32594	hypothetical prote
35	51	40.2	707	2 A46302	PTB-associated spl
36	51	40.2	712	1 I46031	gelatinase B (EC 3
37	51	40.2	1201	2 A83007	hypothetical prote
38	50.5	39.8	157	2 S28012	outB protein - Erw
39	50.5	39.8	377	2 T36246	probable glycolate
40	50.5	39.8	430	2 JC7379	levansucrase (EC 2
41	50	39.4	192	2 E75318	transcription term
42	50	39.4	392	2 T04150	RAD23 protein homo
43	50	39.4	482	2 T36045	hypothetical prote
44	50	39.4	513	2 T03916	hypothetical prote
45	50	39.4	733	2 T03117	glycoprotein H - a

ALIGNMENTS

RESULT 1

A35175  
mucin 1 precursor, repetitive splice form A [validated] - human  
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin  
ncreatic mucin; polymorphic epithelial mucin (PEM)  
N:Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,  
C:Species: Homo sapiens (man)  
C:Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000  
C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51  
R:Digtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilken, J.  
J. Biol. Chem. 265, 5573-5578, 1990  
A:Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc  
A:Reference number: A35175; MUID:90202794; PMID:2318825  
A:Accession: A35175  
A:Molecule type: mRNA  
A:Residues: 1-952,1033-1344 <LIIG1>  
A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124; GB:  
A:Experimental source: splice form A  
A:Note: GenBank entries HUMEPISIA1 and HUMEPISIA2 present only the amino-and carboxyl-er  
A:Accession: B35175  
A:Molecule type: mRNA  
A:Residues: 1-19,29-952,1033-1344 <LIIG2>  
A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129; GB:  
A:Experimental source: splice form B  
A:Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl-en  
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel  
J. Biol. Chem. 265, 15286-15293, 1990  
A:Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli  
A:Reference number: A35886; MUID:90368715; PMID:1697589  
A:Accession: A35886  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-19,29-992,1033-1344 <GEN>  
A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870  
A:Note: GenBank entry HUMUCAB includes one copy of the tandemly repeated sequence  
R:lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.  
J. Biol. Chem. 265, 15294-15299, 1990  
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.  
A:Reference number: A35887; MUID:90368716; PMID:2394722  
A:Accession: A35887  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>  
A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599  
A:Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated seque  
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.  
Eur. J. Biochem. 189, 463-473, 1990  
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera  
A:Reference number: S10571; MUID:90276413; PMID:2351132  
A:Accession: S10572  
A:Molecule type: mRNA  
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>

```

A;Cross-references: EMBL:X52229; NID:g37053
R;Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A;Reference number: S40293
A;Accession: S40293
A;Molecule type: mRNA
A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <NR2>
A;Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
R;Abe, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A;Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A;Reference number: A36735; MUID:90088473; PMID:2597151
A;Accession: A36735
A;Molecule type: mRNA
A;Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A;Cross-references: EMBL:M31823; NID:gl81542; PIDN:AAA35757.1; PID:gl81543
R;Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
J. Biochem. 112, 609-615, 1992
A;Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A;Reference number: JX0235; MUID:93123189; PMID:1478919
A;Accession: PX0066
A;Molecule type: mRNA
A;Residues: 998-1011,'ES',1014-1017;1018-1032,'T',1034-1037;1038-1057 <MAS>
A;Experimental source: Gastric carcinoma cell
R;Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A;Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
A;Reference number: S51026; MUID:95080414; PMID:7988707
A;Contents: annotation
A;Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region an
C;Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C;Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C;Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C;Genetics:
A;Gene: GDB:MUC1; PUM
A;Cross-references: GDB:120705; OMIM:158340
A;Map position: 1q21-1q23
A;Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C;Superfamily: polymorphic epithelial mucin
F;1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
F;1-62/Region: mucin 1 amino-terminal non-repetitive
F;1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
F;1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
F;1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
F;1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
F;138-1317/Region: 20-residue repeats (GSTAPPAHGVTSPADTRPAP)
F;1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F;1245-1272/Domain: transmembrane #status predicted <TRM>
F;1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 127; DB 1; Length 1344;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSPADT 23
|||||
Db 151 PDTRPAPGSTAPPAHGVTSAEET 173

RESULT 2
S10571
mucin 1 precursor, secreted epithelial tumor antigen splice form - human
N;Contains: mucin 1 secreted breast-cancer-associated splice form
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 01-Dec-2000
C;Accession: S10571; JN0100; I56024; S09706; S10217
R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
A;Reference number: S10571; MUID:90276413; PMID:2351132

A;Accession: S10571
A;Molecule type: mRNA
A;Residues: 1-347 <WRE>
A;Cross-references: EMBL:X52228; NID:g36434; PIDN:CAA36477.1; PID:g36435
R;Tsarfaty, I.; Hareuveni, M.; Horev, J.; Zaretsky, J.; Weiss, M.; Jeitsch, J.M.; Garnj
Gene 93, 313-318, 1990
A;Title: Isolation and characterization of an expressed hypervariable gene coding for a
A;Reference number: JN0100; MUID:91033045; PMID:1688329
A;Accession: JN0100
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-135,'Q',137-142,'E',144-163,204-208,'A',210-347 <TSA>
A;Cross-references: GB:M35093; NID:gl82252; PIDN:AAB59612.1; PID:gl82253
R;Xing, P.X.; Tjandra, J.J.; Reynolds, K.; McLaughlin, P.J.; Purcell, D.F.J.; McKenzie,
J. Immunol. 142, 3503-3509, 1989
A;Title: Reactivity of anti-human milk fat globule antibodies with synthetic peptides.
A;Reference number: I56024; MUID:89235154; PMID:2715633
A;Accession: I56024
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 182-201 <RES>
A;Cross-references: GB:M26316; NID:g516622; PIDN:AAA36336.1; PID:g516623
R;Tendler, S.J.B.
Biochem. J. 267, 733-737, 1990
A;Title: Elements of secondary structure in a human epithelial mucin core peptide fragm
A;Reference number: S09706; MUID:90253387; PMID:2339983
A;Accession: S09706
A;Molecule type: protein
A;Residues: 182-201 <TEN>
C;Genetics:
A;Gene: GDB:MUC1; PUM
A;Cross-references: GDB:120705; OMIM:158340
A;Map position: 1q21-1q23
C;Keywords: alternative splicing; tandem repeat
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-347/Product: mucin 1, secreted epithelial tumor antigen splice form #status predic
F;24-163,204-347/Product: mucin 1, secreted breast-cancer-associated splice form #statu

Query Match 96.1%; Score 122; DB 2; Length 347;
Best Local Similarity 95.7%; Pred. No. 1.4e-07;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSPADT 23
|||||
Db 142 PDTRPPPGSTAPPAHGVTSPADT 164

RESULT 3
A60533
tumor-associated antigen DF3 - human
C;Species: Homo sapiens (man)
C;Date: 19-Mar-1993 #sequence_revision 07-May-1993 #text_change 24-Nov-1999
C;Accession: A60533
R;Merlo, G.R.; Siddiqui, J.; Cropp, C.S.; Liscia, D.S.; Lidereau, R.; Callahan, R.; Kuf
Cancer Res. 49, 6966-6971, 1989
A;Title: Frequent alteration of the DF3 tumor-associated antigen gene in primary human
A;Reference number: A60533; MUID:90058554; PMID:2582438
A;Accession: A60533
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-256 <MER>
C;Genetics:
A;Map position: 1q21-q24
C;Superfamily: proline-rich protein
C;Keywords: glycoprotein; tandem repeat

Query Match 81.1%; Score 103; DB 2; Length 256;
Best Local Similarity 78.3%; Pred. No. 2.1e-05;
Matches 18; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSPADT 23
|:::|||||
Db 5 PESRPAPGSTAPPAHRVTSAPES 27

```

## RESULT 4

S52959  
male-specific lethal-1 protein - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 16-Feb-1997  
C;Accession: S52959; S65350  
R;Palmer, M.J.; Mergner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.C.  
Genetics 134, 545-557, 1993  
A;Title: The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes a  
A;Reference number: S52959; MUID:93314941; PMID:8325488  
A;Accession: S52959  
A;Molecule type: DNA  
A;Residues: 1-955 <PAL>  
A;Cross-references: EMBL:L14582  
R;Palmer, M.J.; Mergner, V.A.; Richman, R.; Manning, J.E.; Kuroda, M.I.; Lucchesi, J.C.  
submitted to the EMBL Data Library, April 1993  
A;Reference number: S65350  
A;Accession: S65350  
A;Molecule type: DNA  
A;Residues: 1-407, 'S', 409-585, 'M', 587-955 <PAW>  
A;Cross-references: EMBL:L14582  
C;Genetics:  
A;Gene: msl-1  
A;Cross-references: FlyBase:FBgn0005617  
A;Introns: 329/3

Query Match 48.0%; Score 61; DB 2; Length 955;

Best Local Similarity 59.1%; Pred. No. 8.3;  
Matches 13; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

QY 1 PDTRPAPGSTAPPAHGVTSAPD 22

Db 664 PSTTTPGST--PQHAVTSSMD 683

## RESULT 5

T50568  
probable multi-domain regulatory protein [imported] - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C;Accession: T50568  
R;Redenbach, M.; Kieser, H.M.; Denapaita, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopw  
Mol. Microbiol. 21, 77-96, 1996  
A;Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S  
A;Reference number: Z20556; MUID:97000351; PMID:8843436  
A;Accession: T50568  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1334 <RED>  
A;Cross-references: EMBL:AL133220; PIED:CAB61705.1  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Note: SCC75A.05C

Query Match 48.0%; Score 61; DB 2; Length 1334;

Best Local Similarity 60.9%; Pred. No. 11;  
Matches 14; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

QY 1 PDTRPAPGSTAPPAHGVTSAPDT 23

Db 420 PGTAPAGST--PAPGTVAPGT 440

## RESULT 6

S71753  
repellent protein 1 precursor - smut fungus (Ustilago maydis)  
C;Species: Ustilago maydis (corn smut)  
C;Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 21-Jul-2000  
C;Accession: S71753; S78407  
R;Woesten, H.A.B.; Bohlmann, R.; Eckerskorn, C.; Lottspeich, F.; Boelker, M.; Kahmann, R  
EMBO J. 15, 4274-4281, 1996

A;Title: A novel class of small amphipathic peptides affect aerial hyphal growth and su  
A;Reference number: S71753; MUID:97015128; PMID:8861956

A;Accession: S71753

A;Molecule type: DNA

A;Residues: 1-652 <WOE1>

A;Cross-references: EMBL:U56826; NID:g1336122; PIDN:AAC49419.1; PID:g1336123

A;Experimental source: strain FBD11

A;Accession: S78407

A;Molecule type: prctein

A;Residues: 60-95;154-157;191-195;266-270;313-317;350-354;187-391 <WOE2>

A;Genetics:

A;Gene: rep1

C;Function:

A;Description: plays a structural role in the formation of aerial hyphae

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-59/Domain: propeptide #status predicted <PRO>

F;60-652/Product: repellent protein 1 #status experimental <MAT>

Query Match 46.5%; Score 59; DB 2; Length 652;

Best Local Similarity 47.8%; Pred. No. 10;

Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPDT 23

Db 485 PSTKPEPSPPTTPPSKDTSKPTT 507

## RESULT 7

A71260

hypothetical protein TP0966 - syphilis spirochete

C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C;Accession: A71260

R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwi  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998

A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A;Reference number: A71250; MUID:98332770; PMID:9655876

A;Accession: A71260

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-544 <COL>

A;Cross-references: GB:AE001264; GB:AE000520; NID:g3323278; PIDN:AAC65924.1; PID:g3323232;

A;Experimental source: strain Nichols

C;Genetics:

A;Gene: TP0966

Query Match 44.9%; Score 57; DB 2; Length 544;

Best Local Similarity 62.5%; Pred. No. 15;

Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 PAPGSTAPPAHGVTS A 20

Db 40 PAPGTPAPPAHTASEA 55

## RESULT 8

I52580

gelatinase B (EC 3.4.24.35) precursor - mouse

N;Alternate names: collagenase type IV

C;Species: Mus sp. (mouse)

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C;Accession: I52580

R;Graubert, T.; Johnston, J.; Berliner, N.

Blood 82, 3192-3197, 1993

A;Title: Cloning and expression of the cDNA encoding mouse neutrophil gelatinase: demons

A;Reference number: I52580; MUID:94033534; PMID:8219207

A;Accession: I52580

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-730 <RES>

A;Cross-references: GB:S67830; NID:g460863; PIDN:AAB28942.1; PID:g460864



C;Superfamily: gelatinase A; fibronectin type II repeat homology; hemopexin repeat homol  
C;Keywords: hydrolase; metalloproteinase; zinc; zymogen  
F;230-271/Domain: fibronectin type II repeat homology <2F1>  
F;288-329/Domain: fibronectin type II repeat homology <2F2>  
F;347-388/Domain: fibronectin type II repeat homology <2F3>  
F;529-729/Domain: hemopexin repeat homology <PXN>  
F;100,401,405,411/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status  
F;401,405,411/Binding site: zinc, catalytic (His) (active) #status predicted  
F;402/Active site: Glu #status predicted

Query Match 44.9%; Score 57; DB 1; Length 730;  
Best Local Similarity 47.6%; Pred. No. 19;  
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 21  
| : ||| | : | : |||  
Db 495 PTSSPSPGPTGAPSPGPTAAP 515

RESULT 9  
T50505  
hypothetical protein T22D6.140 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 08-Sep-2000  
C;Accession: T50505  
R;Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De  
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, May 2000  
A;Reference number: Z25101  
A;Accession: T50505  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-407 <BEV>  
A;Cross-references: EMBL:AL357612  
A;Experimental source: cultivar Columbia; BAC clone T22D6  
C;Genetics:  
A;Map position: 5  
A;Introns: 96/3; 140/2  
A;Note: T22D6.140  
C;Superfamily: Arabidopsis thaliana hypothetical protein T22D6.140

Query Match 44.5%; Score 56.5; DB 2; Length 407;  
Best Local Similarity 45.5%; Pred. No. 13;  
Matches 10; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22  
| : ||| | : | : |||  
Db 52 FTSSPPSSSPSPFPG-SNSPD 72

RESULT 10  
S52286  
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain PSST precursor - Arabidopsis thalian  
N;Alternate names: protein T22P22.160  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 08-May-1995 #sequence\_revision 10-Nov-1995 #text\_change 03-Jun-2002  
C;Accession: S52286; T48536  
R;Heiser, V.; Grohmann, L.  
submitted to the EMBL Data Library, January 1995  
A;Reference number: S52286  
A;Accession: S52286  
A;Molecule type: mRNA  
A;Residues: 1-218 <HEI>  
A;Cross-references: EMBL:X84078; NID:G643089; PIDN:CAA58887.1; PID:G643090  
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z24490  
A;Accession: T48536  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-218 <BEV>  
A;Cross-references: EMBL:AL163814  
A;Experimental source: cultivar Columbia; BAC clone T22P22

C;Genetics:  
A;Gene: PSST  
A;Map position: 5  
A;Genome: nuclear  
A;Introns: 141/2  
A;Note: T22P22.160  
C;Superfamily: pbg protein  
C;Keywords: mitochondrion; NAD; oxidoreductase

Query Match 44.1%; Score 56; DB 2; Length 218;  
Best Local Similarity 50.0%; Pred. No. 8.1;  
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 TRPAPGSTAPPAGHGVTSAPD 22  
||| ||| : ||| : :  
Db 47 TRPGPPSTSPPPGGLSKAAE 66

RESULT 11  
JC4364  
gelatinase B (EC 3.4.24.35) precursor - rat  
N;Alternate names: 92-kDa type IV collagenase; matrix metalloproteinase 9 (MMP9)  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 10-Jan-1996 #sequence\_revision 08-Feb-1996 #text\_change 22-Jun-1999  
C;Accession: JC4364  
R;Okada, A.; Santavicca, M.; Bassett, P.  
Gene 164, 317-321, 1995  
A;Title: The cDNA cloning and expression of the gene encoding rat gelatinase B.  
A;Reference number: JC4364; MUID:96069602; PMID:7590350  
A;Accession: JC4364  
A;Molecule type: mRNA  
A;Residues: 1-708 <OKA>  
A;Cross-references: GB:U24441; NID:G1173505; PIDN:AAA90911.1; PID:G1173506  
A;Experimental source: skin wounds  
C;Comment: This enzyme belongs to the matrix metalloproteinase family, proteolytic enzy  
ogression.  
C;Genetics:  
A;Gene: gelB  
C;Superfamily: gelatinase A; fibronectin type II repeat homology; hemopexin repeat hom  
C;Keywords: extracellular matrix; fibroblast; glycoprotein; hydrolase; metalloprotein  
F;1-24/Domain: signal sequence #status predicted <SIG>  
F;25-708/Product: progelatinase B #status predicted <ACT>  
F;25-107/Domain: activation peptide #status predicted <PRO>  
F;108-708/Product: gelatinase B #status predicted <MAT>  
F;214-389/Region: collagen binding #status predicted  
F;231-272/Domain: fibronectin type II repeat homology <2F1>  
F;289-330/Domain: fibronectin type II repeat homology <2F2>  
F;348-389/Domain: fibronectin type II repeat homology <2F3>  
F;514-707/Domain: hemopexin repeat homology <PXN>  
F;39,121/Binding site: carbonylate (Asn) (covalent) #status predicted  
F;100,402,406,412/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #statu  
F;402,406,412/Binding site: zinc, catalytic (His) (active) #status predicted  
F;403/Active site: Glu #status predicted  
F;519-707/Disulfide bonds: #status predicted

Query Match 44.1%; Score 56; DB 2; Length 708;  
Best Local Similarity 47.6%; Pred. No. 25;  
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAP 21  
| : ||| | : | : |||  
Db 488 PTGAPSPGPTGPPTAGPSEAP 508

RESULT 12  
S62907  
gelatinase B (EC 3.4.24.35) precursor - rat  
N;Alternate names: collagenase type IV  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 30-Sep-2001  
C;Accession: S62907; S72371  
R;Xia, Y.; Garcia, G.; Chen, S.; Wilson, C.B.; Feng, L.  
FEBS Lett. 382, 285-288, 1996

A;Title: Cloning of rat 92-kDa type IV collagenase and expression of an active recombinant  
 A;Reference number: S62907; MUID:96184505; PMID:8605986  
 A;Accession: S62907  
 A;Status: preliminary; nucleic acid sequence not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-708 <XIA>  
 A;Cross-references: EMBL:U36476  
 R;Feng, L.  
 submitted to the EMBL Data Library, September 1995  
 A;Reference number: S72371  
 A;Accession: S72371  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-127, 'S', 129-708 <FEN>  
 A;Cross-references: EMBL:U36476; NID:gl022783; PIDN:AAB01721.1; PID:gl022784  
 C;Superfamily: Gelatinase A; fibronectin type II repeat homology; hemopexin repeat homol  
 C;Keywords: extracellular matrix; fibroblast; glycoprotein; hydrolase; metalloproteinase  
 F;1-19/Domain: signal sequence #status predicted <SIG>  
 F;20-107/Domain: propeptide #status predicted <PRO>  
 F;25-107/Domain: activation peptide #status predicted <ACT>  
 F;68-217,393-445/Domain: matrix metalloproteinase homology #status atypical <MMP>  
 F;128-708/Product: gelatinase A #status predicted <MAT>  
 F;214-389/Region: collagen binding #status predicted  
 F;231-272/Domain: fibronectin type II repeat homology <2F1>  
 F;289-330/Domain: fibronectin type II repeat homology <2F2>  
 F;348-389/Domain: fibronectin type II repeat homology <2F3>  
 F;514-707/Domain: hemopexin repeat homology <PXN>  
 F;100,402,406,412/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status  
 F;402,406,412/Binding site: zinc, catalytic (His) (active) #status predicted  
 F;403/Active site: Glu #status predicted  
 F;519-707/Disulfide bonds: #status predicted

Query Match 44.1%; Score 56; DB 2; Length 708;

Best Local Similarity 47.6%; Pred. No. 25;

Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAP 21

Db 488 FTGAPSPGPTGPTAGPSEAP 508

# RESULT 13

A11347  
 hypothetical protein lmo2185 [imported] - Listeria monocytogenes (strain EGD-e)  
 C;Species: Listeria monocytogenes  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: A11347  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
 O.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
 ck, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A;Title: Comparative genomics of Listeria species.  
 A;Reference number: AB1077; MUID:21537279; PMID:11679669  
 A;Accession: A11347  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-569 <GLA>  
 A;Cross-references: GB:NC\_003210; PIDN:CAD00263.1; PID:gl16411655; GSPDB:GN00177  
 A;Experimental source: strain EGD-e  
 C;Genetics:  
 A;Gene: lmo2185

Query Match 43.7%; Score 55.5; DB 2; Length 569;

Best Local Similarity 52.2%; Pred. No. 23;

Matches 12; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 1 PDTRPAPGSTAPPAGHGVTSAPPT 23

Db 328 PVTKPDGPTTNP-----VTTPTPT 347

## RESULT 14

A34458  
 gelatinase B (EC 3.4.24.35) precursor - human  
 N;Alternate names: 92K gelatinase; 92K type IV collagenase; matrix metalloproteinase 9  
 N;Contains: 91K neutrophil gelatinase B  
 C;Species: Homo sapiens (man)  
 C;Date: 03-Aug-1992 #sequence\_revision 07-Apr-1994 #text\_change 18-Jun-1999  
 C;Accession: A34458; A41166; A42253; A45114; B48417; A61385; S16097; S59488  
 R;Wilhelm, S.M.; Collier, I.E.; Marmer, B.L.; Eisen, A.Z.; Grant, G.A.; Goldberg, G.I.  
 J. Biol. Chem. 264, 17213-17221, 1989  
 A;Title: SV40-transformed human lung fibroblasts secrete a 92-kDa type IV collagenase w/  
 A;Reference number: A34458; MUID:90008879; PMID:2551898  
 A;Accession: A34458  
 A;Molecule type: mRNA  
 A;Residues: 1-11, 'L', 13-667, 'R', 669-707 <WIL>  
 A;Cross-references: GB:J05070; NID:gl177204; PIDN:AAA51539.1; PID:gl177205  
 R;Huhtala, P.; Tuuttila, A.; Chow, L.T.; Lohi, J.; Keski-Oja, J.; Tryggvason, K.  
 J. Biol. Chem. 266, 16485-16490, 1991  
 A;Title: Complete structure of the human gene for 92-kDa type IV collagenase. Divergent  
 A;Reference number: A41166; MUID:91358433; PMID:1653238  
 A;Accession: A41166  
 A;Molecule type: DNA  
 A;Residues: 1-49, 121-127, 171-177, 214-220, 272-278, 330-336, 389-395, 439-542, 582-587, 631-637  
 A;Cross-references: GB:M68343; GB:M68344; GB:M68345; GB:M68346; GB:M68347; GB:M68348; GI  
 A;Note: the authors translated the codon CTC for residue 276 as Lys  
 R;Ogata, Y.; Enghild, J.J.; Nagase, H.  
 J. Biol. Chem. 267, 3581-3584, 1992  
 A;Title: Matrix metalloproteinase 3 (stromelysin) activates the precursor for the human  
 A;Reference number: A42253; MUID:92156083; PMID:1371271  
 A;Accession: A42253  
 A;Molecule type: protein  
 A;Residues: 20-34, 60-71, 107-118 <OGA>  
 R;Okada, Y.; Gonoji, Y.; Naka, K.; Tomita, K.; Nakanishi, I.; Iwata, K.; Yamashita, K.;  
 J. Biol. Chem. 267, 21712-21719, 1992  
 A;Title: Matrix metalloproteinase 9 (92-kDa gelatinase/type IV collagenase) from HT 1080  
 A;Reference number: A45114; MUID:93016125; PMID:1400481  
 A;Accession: A45114  
 A;Molecule type: protein  
 A;Residues: 20-32 <OKA>  
 A;Experimental source: HT 1080 fibrosarcoma cells  
 A;Note: sequence extracted from NCBI backbone (NCBIP:119576)  
 A;Accession: B45114  
 A;Molecule type: protein  
 A;Residues: 94-111 <OK2>  
 A;Experimental source: HT 1080 fibrosarcoma cells  
 A;Note: sequence extracted from NCBI backbone (NCBIP:116742)  
 A;Note: amino ends of the proenzyme and two activated forms  
 R;Van Ranst, M.; Norga, K.; Masure, S.; Proost, P.; Vandekerckhove, F.; Auwerx, J.; Van  
 Cytokine 3, 231-239, 1991  
 A;Title: The cytokine-protease connection: identification of a 96-kD THP-1 gelatinase ar  
 A;Reference number: A48417; MUID:91355647; PMID:1653055  
 A;Accession: B48417  
 A;Molecule type: protein  
 A;Residues: 20-25, 'X', 27-37 <VAN>  
 A;Experimental source: monocytic cell line THP-1  
 A;Note: sequence extracted from NCBI backbone (NCBIP:63812) and corrected to correspond  
 R;Opdenakker, G.; Masure, S.; Grillet, B.; Van Damme, J.  
 Lymphokine Cytokine Res. 10, 317-324, 1991  
 A;Title: Cytokine-mediated regulation of human leukocyte gelatinases and role in arthrit  
 A;Reference number: A61385; MUID:92032113; PMID:1932376  
 A;Accession: A61385  
 A;Molecule type: protein  
 A;Residues: 28-37 <CPD>  
 R;Masure, S.; Proost, P.; van Damme, J.; Opdenakker, G.  
 Eur. J. Biochem. 198, 391-398, 1991  
 A;Title: Purification and identification of 91-kDa neutrophil gelatinase. Release by the  
 A;Reference number: S16097; MUID:91249834; PMID:1645657  
 A;Accession: S16097  
 A;Molecule type: protein  
 A;Residues: 28-42, 'X', 44-60, 'XX', 63 <MAS>  
 R;Sang, Q.X.; Birkedal-Hansen, H.; van Wart, H.E.  
 Biochim. Biophys. Acta 1251, 99-108, 1995  
 A;Title: Proteolytic and non-proteolytic activation of human neutrophil progelatinase B.



C;Keywords: extracellular matrix; fibroblast; glycoprotein; hydrolase; metalloproteinase  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-730/Product: progelatinase A #status predicted <PRO>  
F:25-107/Domain: activation peptide #status predicted <ACT>  
F:108-730/Product: gelatinase A #status predicted <MAT>  
F:213-388/Region: collagen binding #status predicted  
F:230-271/Domain: fibronectin type II repeat homology <2F1>  
F:288-329/Domain: fibronectin type II repeat homology <2F2>  
F:347-388/Domain: fibronectin type II repeat homology <2F3>  
F:453-549/Region: PEST sequence  
F:529-729/Domain: hemopexin repeat homology <PXN>  
F:100,401,405,411/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status  
F:401,405,411/Binding site: zinc, catalytic (His) (active) #status predicted  
F:402/Active site: Glu #status predicted  
F:534-729/Disulfide bonds: #status predicted

Query Match 42.5%; Score 54; DB 2; Length 730;  
Best Local Similarity 47.6%; Pred. NO. 45;  
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

```

QY      1 PDTRPAGSTAPPAHGVTSAP 21
      | | | | | | | | | |
Db     487 PTGAPSPGPTSSPSPGPTCAP 507

```

## RESULT 18

T00357  
hypothetical protein KIAA0685 - human  
C/Species: Homo sapiens (man)  
C/Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C/Accession: T00357  
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, M.  
DNA Res. 5, 169-176, 1998  
A/Title: Prediction of the coding sequences of unidentified human genes. X. The complete  
A/Reference number: Z14142; MUID:98403880; PMID:9734811  
A/Accession: T00357  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-927 <ISH>  
A/Cross-references: EMBL:AB014585; NID:g3327183; PIDN:BAA31660.1; PID:g3327184  
A/Experimental source: brain  
C/Genetics:  
A/Note: KIAA0685

Query Match 42.5%; Score 54; DB 2; Length 927;  
Best Local Similarity 61.1%; Pred. NO. 56;  
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 RPAPGSTAPFAHGVTSA 21  
Db 658 RDAPGAGAPFAPGKKEAP 675

RESULT 19

F72570  
hypothetical protein APE1847 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
C;Accession: F72570  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1993  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: F72570  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-115 <KAW>  
A;Cross-references: DDBJ:AP000062; NID:G5105244; PIDW:BAA80851.1; PID:d1044637; PID:G510 A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE1847  
C;Superfamily: Aeropyrum pernix hypothetical protein APE1847

Query Match 40.9%; Score 52; DB 2; Length 115;  
Best Local Similarity 50.0%; Pred. NO. 13;  
Matches 11; Conservative 4; Mismatches 3; Indels

Qy	3	TRPAPGSTAPPA---HGV TSA	20
		:       :	:
Dd	88	TRSPGSGTGPSTLSRHGLSWA	109

RESULT 20

T51838  
blue copper binding protein homolog [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 21-Oct-2002  
C:Accession: T51838  
R:Yang, K.Y.; Kim, C.S.; Cho, B.H.  
submitted to the EMBL Data Library, October 1998  
A:Description: Characterization of a wound-inducible Arabidopsis gene encoding a protein  
A:Reference number: Z25481  
A:Accession: T51838  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-196 <VAN>  
A:Cross-references: EMBL:Y18227; PIDN:CAA77089.1  
C:Genetics:  
A:Gene: AWI 32  
C:Superfamily: stellacyanin  
C:Keywords: copper binding

Query Match 40.9%; Score 52; DB 2; Length 196;  
Best Local Similarity 42.9%; Pred. No. 22;  
Matches 9; Conservative 3; Mismatches 9; Indels

QY	1	P	D	T	R	P	A	G	T	S	A	P	21
			:		:					:			
Db	142	P	G	S	T								

RESULT 21  
A49933

proteinase SM transporter PrtD - *Serratia marcescens*  
N;Alternate names: inner membrane metalloprotease transporter PrtD  
C;Species: *Serratia marcescens*  
C;Date: 13-Sep-1994 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001  
C;Accession: A49933  
R;Letoffe, S.; Ghigo, J.M.; Wandersman, C.  
J. Bacteriol. 175, 7321-7328, 1993  
A;Title: Identification of two components of the *Serratia marcescens* metalloprotease tra  
A;Reference number: A49933; MUID:94042907; PMID:8226679  
A;Contents: 365

Query Match 40.9%; Score 52; DB 2; Length 576;  
Best Local Similarity 62.5%; Pred. NO. 62;  
Matches 10; Conservative 1; Mismatches 5; Indels

QY 3 TRPAPGSGTAPPAHGVT 18  
| ||| ||| ||| :  
Db 328 TASAPGSGTAPVLHGVS 34

RESULT 22  
T48672

T48672



ABC-type transport system ATP-binding protein hasD [validated] - Serratia marcescens  
N;Alternate names: hasA export system ABC-type transport component hasD  
N;Contains: adenosinetriphosphatase (EC 3.6.1.3)  
C;Species: Serratia marcescens  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Sep-2000  
C;Accession: T48672  
R;Letoffe, S.; Ghigo, J.M.; Wandersman, C.  
Proc. Natl. Acad. Sci. U.S.A. 91, 9876-9880, 1994  
A;Title: Iron acquisition from heme and hemoglobin by a Serratia marcescens extracellular  
A;Reference number: Z24524; MUID:95024064; PMID:7937909  
A;Accession: T48672  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-586 <LET>  
A;Cross-references: EMBL:X81195; PIDN:CAA57069.1  
A;Experimental source: strain SM365  
C;Genetics:  
A;Gene: hasD  
C;Complex: the hasA export system consists of the ATP-binding protein hasD (PIR:T48672),  
D:97368101  
C;Function: <ATP>  
A;Description: EC 3.6.1.3  
C;Function: <HAS>  
A;Description: involved in the export of the heme acquisition protein hasA [validated, M  
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
C;Keywords: ATP binding; hydrolase; membrane-associated protein; protein export

Query Match 40.9%; Score 52; DB 2; Length 586;  
Best Local Similarity 62.5%; Pred. No. 63;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 TRPAPGSTAPPAHGVGT 18  
| ||| ||| |||  
Db 338 TASAPGCTAPVLHGVS 353

RESULT 23  
A39344  
tumor-associated mucin (MUC1) homolog precursor - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 03-Apr-1992 #sequence\_revision 03-Apr-1992 #text\_change 23-Feb-1997  
C;Accession: A39344  
R;Spicer, A.P.; Parry, G.; Patton, S.; Gendler, S.J.  
J. Biol. Chem. 266, 15099-15109, 1991  
A;Title: Molecular cloning and analysis of the mouse homologue of the tumor-associated m  
a loss of minisatellite-like polymorphism.  
A;Reference number: A39344; MUID:91332029; PMID:1714452  
A;Accession: A39344  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-630 <SPI>  
A;Cross-references: GB:M64928  
C;Keywords: cytoskeleton; transmembrane protein

Query Match 40.9%; Score 52; DB 2; Length 630;  
Best Local Similarity 47.8%; Pred. No. 67;  
Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAHGVTSAPDT 23  
| ||| ||| ||| |||  
Db 55 PATRPPGDSSTSSPVQSSSTSPAT 77

RESULT 24  
I52257  
episialin - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 05-Nov-1999  
C;Accession: I52257; I65210  
R;Vos, H.L.; De Vries, Y.; Hilkens, J.  
Biochem. Biophys. Res. Commun. 181, 121-130, 1991  
A;Title: The mouse episialin (Muc1) gene and its promoter. Rapid evolution of the repeti  
A;Reference number: I52257; MUID:92068178; PMID:1958179

A;Accession: I52257  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-631 <RES>  
A;Cross-references: GB:M77226; NID:g199835; PIDN:AAA39754.1; PID:g199837  
A;Accession: I65210  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-631 <RE2>  
A;Cross-references: GB:M84683; NID:g199842; PIDN:AAA39756.1; PID:g199843  
C;Genetics:  
A;Gene: Muc1  
A;Introns: 20/1; 454/3; 472/2; 517/1; 557/3; 607/3

Query Match 40.9%; Score 52; DB 2; Length 631;  
Best Local Similarity 47.8%; Pred. No. 68;  
Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAHGVTSAPDT 23  
| ||| ||| ||| |||  
Db 55 PATRPPGDSSTSSPVQSSSTSPAT 77

RESULT 25  
B97566  
hypothetical protein AGR\_C\_3142 [imported] - Agrobacterium tumefaciens (strain C58, Ce  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C;Accession: B97566  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldmar  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium t  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: B97566  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-912 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK87483.1; PID:g15156806; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C\_3142  
A;Map position: circular chromosome

Query Match 40.9%; Score 52; DB 2; Length 912;  
Best Local Similarity 52.6%; Pred. No. 96;  
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 DTRPAPGSTAPPAHGVTS 20  
: ||| | ||| | :  
Db 148 EAAPAPVSVRPPAHGGRSS 166

RESULT 26  
AH2786  
conserved hypothetical protein Atul710 [imported] - Agrobacterium tumefaciens (strain (C  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AH2786  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCle  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193  
A;Accession: AH2786  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1008 <KUR>  
A;Cross-references: GB:AE008688; PIDN:AAL42710.1; PID:g17740147; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:



A;Gene: Atul710  
A;Map position: circular chromosome

Query Match 40.9%; Score 52; DB 2; Length 1008;  
Best Local Similarity 52.6%; Pred. No. 1.1e+02;  
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 DTRPAPGSTAPPAHGVTSAPDT 20  
Db 244 EAAPAPVSVRPPPAHGGRSS 262

RESULT 27

T28155  
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragment  
N;Alternate names: erythrocyte membrane binding protein 1 (EMPI)  
C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C;Accession: T28155  
R;Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.  
Nature 388, 292-295, 1997  
A;Title: Plasmodium falciparum rosetting is mediated by PfEMP1 and requires complement  
A;Reference number: T20477; MUID:97373957; PMID:9230440  
A;Accession: T28155  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2706 <ROW>  
A;Cross-references: EMBL:Y13402; PIDN:CAA73831.1  
A;Experimental source: strain IT 4/25/5  
C;Genetics:  
A;Introns: 2493/3  
A;Note: R29R+var1

Query Match 40.9%; Score 52; DB 2; Length 2706;  
Best Local Similarity 64.3%; Pred. No. 2.7e+02;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPA 14  
Db 2450 PDTRPVPSPLPPA 2463

RESULT 28

T13423  
hypothetical protein 30B8.4 - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C;Accession: T13423  
R;Murphy, L.; Harris, D.; Barrell, B.  
submitted to the EMBL Data Library, April 1999  
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A;Reference number: Z17668  
A;Accession: T13423  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-3345 <MUR>  
A;Cross-references: EMBL:AL009195; NID:e1355203; PID:e1248585; PIDN:CAA15708.1  
C;Genetics:  
A;Cross-references: FlyBase:FBgn0000377  
A;Map position: X  
A;Introns: 51/3; 159/1; 476/1; 526/1; 1465/1; 1826/3; 1947/3; 2081/1; 2196/3; 3007/3  
A;Note: EG:30B8.4

Query Match 40.9%; Score 52; DB 2; Length 3345;  
Best Local Similarity 52.4%; Pred. No. 3.3e+02;  
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 3 TRPAPGSTAPPAHGVTSAPDT 23  
Db 339 TNKARGQTNPRQHFTSAPST 359

RESULT 29

A48725  
nuclear DNA-binding protein - Marek's disease virus  
C;Species: Marek's disease virus  
C;Date: 16-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
C;Accession: A48725  
R;Camp, H.S.; Silva, R.F.; Coussens, P.M.  
Virology 196, 484-495, 1993  
A;Title: Defective Marek's disease virus DNA contains a gene encoding a potential nucle  
A;Reference number: A48725; MUID:93383373; PMID:8396799  
A;Accession: A48725  
A;Status: preliminary  
A;Molecule type: nucleic acid  
A;Residues: 1-204 <CAM>  
A;Cross-references: GB:L10087; NID:g306312; PIDN:AAA46121.1; PID:g306313  
A;Experimental source: serotype 2, 281M1/1  
A;Note: sequence extracted from NCBI backbone (NCBIN:137884, NCBIP:137885)

Query Match 40.6%; Score 51.5; DB 2; Length 204;  
Best Local Similarity 71.4%; Pred. No. 26;  
Matches 10; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 4 RPAPG-STAPPAHG 16  
Db 9 RPAPG-ETPPPSHG 22

RESULT 30

S28444  
tonB protein - Pseudomonas putida  
C;Species: Pseudomonas putida  
C;Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 24-Nov-1999  
C;Accession: S28444  
R;Bitter, W.; Tommaseen, J.; Weisbeek, P.J.  
Mol. Microbiol. 7, 117-130, 1993  
A;Title: Identification and characterization of the exbB, exbD and tonB genes of Pseudo  
A;Reference number: S28442; MUID:93172953; PMID:8437515  
A;Accession: S28444  
A;Molecule type: DNA  
A;Residues: 1-243 <BIT>  
A;Cross-references: EMBL:X70139; NID:g49192; PIDN:CAA49716.1; PID:g49195  
C;Genetics:  
A;Gene: tonB  
C;Superfamily: tonB protein

Query Match 40.2%; Score 51; DB 2; Length 243;  
Best Local Similarity 42.9%; Pred. No. 36;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAP 21  
Db 104 PEKXPPPPKEAPPTTEVVDP 124

RESULT 31

T16486  
hypothetical protein F56D3.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C;Accession: T16486  
R;Wilcox, L.  
submitted to the EMBL Data Library, June 1995  
A;Description: The sequence of C. elegans cosmid F56D3.  
A;Reference number: Z18520  
A;Accession: T16486  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-374 <WIL>  
A;Cross-references: EMBL:U28942; NID:g861330; PID:g861332; PIDN:AAA68355.1; CESP:F56D3.1  
A;Experimental source: strain Bristol N2  
C;Genetics:  
A;Gene: CESP:F56D3.1  
A;Introns: 21/2; 41/1; 121/2; 150/2; 245/2; 315/3; 338/3

```
Query Match      40.2%; Score 51; DB 2; Length 374;
Best Local Similarity 57.9%; Pred. No. 54;
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY      3  TTPAGSTAPPAHGVTSAP 21
Db      136  TTPATFATLPPAPAVTRAP 154

RESULT 32
C75530
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: C75530
R;White, O.; Eisen, J.A.; Heidelberg, J.P.; Hickey, B.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: C75530
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-571 <WHI>
A;Cross-references: GB:AE001895; GB:AE000513; NID:96458024; PIDN:AAF09932.1; PID:9645802
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0350
A;Map position: 1

Query Match      40.2%; Score 51; DB 2; Length 571;
Best Local Similarity 55.6%; Pred. No. 81;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY      1  PDTRPAPGSTAPPAHGVTT 18
Db      481  PGTTAPDTPAAPPVAVDTT 498

RESULT 33
T12995
pectinesterase homolog T21L8.150 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 22-Oct-1999
C;Accession: T12995
R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
submitted to the Protein Sequence Database, July 1999
A;Reference number: Z17586
A;Accession: T12995
A;Molecule type: DNA
A;Residues: 1-594 <CHO>
A;Cross-references: EMBL:AL096860; GSPDB:GN00061; ATSP:T21L8.150
A;Experimental source: cultivar Columbia; BAC clone T21L8
C;Genetics:
A;Gene: ATSP:T21L8.150
A;Map position: 3
A;Introns: 251/1; 363/1
C;Superfamily: pectinesterase

Query Match      40.2%; Score 51; DB 2; Length 594;
Best Local Similarity 37.0%; Pred. No. 84;
Matches 10; Conservative 6; Mismatches 5; Indels 6; Gaps 1;

QY      1  PDTRP-----APGSTAPPAHGVTSAP 21
Db      46  PETAPSPTPTSSPSTTSPSPGPVAP 72

RESULT 34
T32594
hypothetical protein C02B10.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
```

```
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32594
R;Nelson, J.; White, S.; Hawkins, J.; Wohldmann, P.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid C02B10.
A;Reference number: Z21196
A;Accession: T32594
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-698 <NEL>
A;Cross-references: EMBL:AF038605; PIDN:AAB92020.1; GSPDB:GN00022; CESP:C02B10.5
A;Experimental source: strain Bristol N2; clone C02B10
C;Genetics:
A;Gene: CESP:C02B10.5
A;Map position: 4
A;Introns: 61/3; 102/2; 188/3; 349/2; 641/1

Query Match      40.2%; Score 51; DB 2; Length 698;
Best Local Similarity 42.9%; Pred. No. 98;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY      1  PDTRPAPGSTAPPAHGVTSAP 21
Db      62  PEPSPAHGRSAPSSQETITSP 82

RESULT 35
A46302
PTB-associated splicing factor, long form - human
N;Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding 1
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A46302; A43557; S29995
R;Patton, J.G.; Porro, E.B.; Galceran, J.; Tempst, P.; Nadal-Ginard, B.
Genes Dev. 7, 393-406, 1993
A;Title: Cloning and characterization of PSF, a novel pre-mRNA splicing factor.
A;Reference number: A46302; MUID:93194059; PMID:8449401
A;Accession: A46302
A;Molecule type: mRNA
A;Residues: 1-707 <PAT>
A;Cross-references: EMBL:X70944; NID:G38457; PIDN:CAA50283.1; PID:G38458
A;Note: sequence extracted from NCBI backbone (NCBIP:127206)
R;Gower, H.J.; Moore, S.E.; Dickson, G.; Elsom, V.L.; Nayak, R.; Walsh, F.S.
Development 105, 723-731, 1989
A;Title: Cloning and characterization of a myoblast cell surface antigen defined by 24
A;Reference number: A43557; MUID:90091812; PMID:2480877
A;Accession: A43557
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 312-707 <GOW>
A;Cross-references: GB:X16850; NID:G23711; PIDN:CAA34747.1; PID:G23712
C;Genetics:
A;Gene: GDB:SFPQ; PSF
A;Cross-references: GDB:138275
A;Map position: 4q-4q
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprot
C;Keywords: alternative splicing; pre-mRNA splicing; surface antigen
F;298-359/Domain: ribonucleoprotein repeat homology <RRM1>
F;372-438/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match      40.2%; Score 51; DB 2; Length 707;
Best Local Similarity 57.1%; Pred. No. 99;
Matches 12; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

QY      1  PDTRPAPGSTAPPAHGVTSAP 21
Db      143  PGSGPGTPTPPPA--VTSAP 161

RESULT 36
I46031
gelatinase B (EC 3.4.24.35) - bovine
N;Alternate names: matrix metalloproteinase 9 (MMP9)
```



Db 376 DTLPARGSTDPQNPAAYRIGGTLP 400

Search completed: May 6, 2004, 16:31:53  
Job time : 15.9231 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:22:19 ; Search time 12.3846 Seconds  
(without alignments)  
96.702 Million cell updates/sec

Title: US-10-070-566-7  
Perfect score: 127  
Sequence: 1 PDTRPAPGSTAPPAGHGVTSAPDT 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	100.0	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
2	113	89.0	475	1 MUC1_HYLLA	Q29435 hylobates l
3	61	48.0	1039	1 MSL1_DROME	P50535 drosophila
4	58	45.7	704	1 MM09_CANFA	O18733 canis famil
5	57	44.9	544	1 Y966_TREPA	O83932 treponema p
6	57	44.9	730	1 MM09_MOUSE	P41245 mus musculu
7	56	44.1	218	1 NUKM_ARATH	Q42577 arabidopsis
8	56	44.1	708	1 MM09_RAT	P50282 rattus norv
9	55	43.3	707	1 MM09_HUMAN	P14780 homo sapien
10	54	42.5	564	1 ZYX_MOUSE	Q62523 mus musculu
11	54	42.5	927	1 K685_HUMAN	O75170 homo sapien
12	54	42.5	1856	1 MGA_HUMAN	O43451 homo sapien
13	53	41.7	1575	1 SYJ1_HUMAN	O43426 homo sapien
14	52	40.9	196	1 BCB1_ARATH	Q07488 arabidopsis
15	52	40.9	630	1 MUC1_MOUSE	Q02496 mus musculu
16	51	40.2	243	1 TONB_PSEPU	Q05613 pseudomonas
17	51	40.2	707	1 SFPQ_HUMAN	P23246 homo sapien
18	51	40.2	712	1 MM09_BOVIN	P52176 bos taurus
19	50.5	39.8	220	1 GSPB_ERWCH	O01563 erwinia chr
20	50.5	39.8	263	1 VGLB_BRSVW	O10687 bovine resp
21	50	39.4	192	1 NUSB_DEIRA	Q91847 deinococcus
22	50	39.4	628	1 ASH2_HUMAN	Q9ub13 homo sapien
23	50	39.4	686	1 KITH_HSVEM	P33802 bovine herp
24	50	39.4	889	1 IF2_RHIME	Q928w4 rhizobium m
25	50	39.4	1229	1 P121_HUMAN	Q9y2n3 homo sapien
26	49.5	39.0	259	1 YL14_HUMAN	P40222 homo sapien
27	49.5	39.0	776	1 RTN1_HUMAN	Q16799 homo sapien
28	49	38.6	118	1 MOBC_THIFE	P22899 thibacillu
29	49	38.6	303	1 HYPB_AZOVI	P31880 azotobacter
30	49	38.6	356	1 HXB2_HUMAN	P14652 homo sapien
31	49	38.6	381	1 APA_MYCAV	Q48919 mycobacteri
32	49	38.6	383	1 VGLZ_HSVBK	P32515 equine herp
33	49	38.6	497	1 HMES_DROME	P18488 drosophila

34	49	38.6	682	1 TUPI_KLULA	P56094 kluyveromyc
35	49	38.6	753	1 CATE_ECOLI	P21179 escherichia
36	49	38.6	776	1 CG2F_MOUSE	P51944 mus musculu
37	49	38.6	1255	1 DIA1_MOUSE	Q08808 mus musculu
38	48.5	38.2	333	1 NK32_HUMAN	P78367 homo sapien
39	48.5	38.2	893	1 DAG1_MOUSE	Q62165 mus musculu
40	48.5	38.2	1033	1 IF2_STRCO	Q8cjq8 streptomyce
41	48.5	38.2	1327	1 TNK1_HUMAN	O95271 homo sapien
42	48.5	38.2	2142	1 BAT2_HUMAN	P48634 homo sapien
43	48	37.8	131	1 C79A_LOEMI	P45586 locusta mig
44	48	37.8	131	1 C79B_LOEMI	P45587 locusta mig
45	48	37.8	156	1 BCCP_PSEAE	P37799 pseudomonas

ALIGNMENTS

RESULT 1					
MUC1_HUMAN					
ID	MUC1_HUMAN	STANDARD:	PRT;	1255 AA.	
AC	P15941; P13931; P15942; P17626; Q14128; Q14876; Q16437; Q16442;				
AC	Q16615; Q9BXA4; Q9UE75; Q9UE76; Q9UQL1; Q9Y4J2;				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-APR-1990 (Rel. 14, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)				
DE	(Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)				
DE	(Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-				
DE	reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen				
DE	DF3) (CD227 antigen).				
GN	MUC1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RC	TISSUE=Pancreas;				
RX	MEDLINE=90368716; PubMed=2394722;				
RA	Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;				
RT	"Cloning and sequencing of a human pancreatic tumor mucin cDNA.";				
RL	J. Biol. Chem. 265:15294-15299(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).				
RX	MEDLINE=90202794; PubMed=2318825;				
RA	Ligtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;				
RT	"Episialin, a carcinoma-associated mucin, is generated by a				
RT	polymorphic gene encoding splice variants with alternative amino				
RT	termini.";				
RL	J. Biol. Chem. 265:5573-5578(1990).				
RN	[3]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RC	TISSUE=Breast carcinoma;				
RX	MEDLINE=90368715; PubMed=1697589;				
RA	Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,				
RA	Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;				
RT	"Molecular cloning and expression of human tumor-associated				
RT	polymorphic epithelial mucin.";				
RL	J. Biol. Chem. 265:15286-15293(1990).				
RN	[4]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RX	MEDLINE=91097524; PubMed=2268309;				
RA	Lancaster C.A., Peat N., Duhig T., Wilson D.,				
RA	Taylor-Papadimitriou J., Gendler S.J.;				
RT	"Structure and expression of the human polymorphic epithelial mucin				
RT	gene: an expressed VNTR unit.";				
RL	Biochem. Biophys. Res. Commun. 173:1019-1029(1990).				
RN	[5]				
RP	SEQUENCE FROM N.A. (ISOFORM 5).				
RC	TISSUE=Breast carcinoma;				
RX	MEDLINE=90276413; PubMed=2351132;				
RA	Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,				
RA	Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;				



RT "Human epithelial tumor antigen cDNA sequences. Differential splicing  
RT may generate multiple protein forms.";  
RL Eur. J. Biochem. 189:463-473(1990).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=90276414; PubMed=2112460;  
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J., Wreschner D.H.;  
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;  
RA "A transcribed gene, containing a variable number of tandem repeats,  
RT codes for a human epithelial tumor antigen. cDNA cloning, expression  
RT of the transfected gene and over-expression in breast cancer  
RT tissue.";  
RL Eur. J. Biochem. 189:475-486(1990).  
RN [7]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=91033045; PubMed=1688329;  
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,  
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;  
RA "Isolation and characterization of an expressed hypervariable gene  
RT coding for a breast-cancer-associated antigen.";  
RL Gene 93:313-318(1990).  
RN [8]  
RP SEQUENCE FROM N.A. (ISOFORM 7).  
RX MEDLINE=95010060; PubMed=7925397;  
RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,  
RA Keydar I., Hilkens J., Wreschner D.H.;  
RA "Characterization and molecular cloning of a novel MUC1 protein,  
RT devoid of tandem repeats, expressed in human breast cancer tissue.";  
RL Eur. J. Biochem. 224:787-795(1994).  
RN [9]  
RP SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).  
RX MEDLINE=97355747; PubMed=9212228;  
RA Oosterkamp H.M., Scheiner L., Stefanowa M.C., Lloyd K.O.,  
RA Finstad C.L.;  
RA "Comparison of MUC-1 mucin expression in epithelial and non-epithelial  
RT cancer cell lines and demonstration of a new short variant form  
RT (MUC-1/Z)." ;  
RL Int. J. Cancer 72:87-94(1997).  
RN [10]  
RP SEQUENCE FROM N.A. (ISOFORM 7).  
RX Zhang L.X., Li C.H.;  
RA "Molecular cloning of an isoform of MUC1, MUC1/Y." ;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [11]  
RP SEQUENCE FROM N.A. (ISOFORM 9).  
RC TISSUE=Epithelial cancer;  
RA Zhang L.X., Li C.H., Sun L.Y., Yue W.;  
RA "Cloning of a new potential secreted short variant form of MUC1 mucin  
RT in epithelial cancer cell line." ;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [12]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=88330762; PubMed=3417635;  
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,  
RA Burchell J.;  
RA "A highly immunogenic region of a human polymorphic epithelial mucin  
RT expressed by carcinomas is made up of tandem repeats." ;  
RL J. Biol. Chem. 263:12820-12823(1988).  
RN [13]  
RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM 2).  
RX MEDLINE=90088473; PubMed=2597151;  
RA Abe M., Siddiqui J., Kufe D.;  
RA "Sequence analysis of the 5' region of the human DF3 breast  
RT carcinoma-associated antigen gene." ;  
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).  
RN [14]  
RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).  
RC TISSUE=Thyroid;  
RX MEDLINE=96183746; PubMed=8608966;  
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;  
RA "Preoperative diagnosis of thyroid papillary carcinoma by reverse  
RT transcriptase polymerase chain reaction of the MUC1 gene." ;

RL Int. J. Cancer 66:55-59(1996).  
RN [15]  
RP SEQUENCE OF 1-89 FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=96181716; PubMed=8604237;  
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,  
RA Lee I.N., Luh K.T., Wu C.W.;  
RA "Mucin mRNA expression in lung adenocarcinoma cell lines and  
RT tissues." ;  
RL Oncology 53:118-126(1996).  
RN [16]  
RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).  
RC TISSUE=Breast carcinoma;  
RA Buluwela L., Liu Q., Iugmani Y.A., Gomm J.J., Coombes R.C.;  
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
RN [17]  
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.  
RX MEDLINE=97460054; PubMed=9312074;  
RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,  
RA Hanisch F.-G.;  
RA "Localization of O-glycosylation sites on glycopeptide fragments from  
RT lactation-associated MUC1. All putative sites within the tandem  
RT repeat are glycosylation targets in vivo." ;  
RL J. Biol. Chem. 272:24780-24793(1997).  
RN [18]  
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.  
RX MEDLINE=99303572; PubMed=10373415;  
RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,  
RA Hanisch F.-G.;  
RA "High density O-glycosylation on tandem repeat peptide from secretory  
RT MUC1 of T47D breast cancer cells." ;  
RL J. Biol. Chem. 274:18165-18172(1999).  
RN [19]  
RP POLYMORPHISM WITHIN THE REPEAT.  
RX MEDLINE=21359366; PubMed=11350974;  
RA Engelmann K., Baldus S.B., Hanisch F.-G.;  
RA "Identification and topology of variant sequences within individual  
RT repeat domains of the human epithelial tumor mucin MUC1." ;  
RL J. Biol. Chem. 276:27764-27769(2001).  
RN [20]  
RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.  
RX MEDLINE=99211485; PubMed=10197628;  
RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,  
RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,  
RA Wreschner D.H.;  
RA "The breast cancer-associated MUC1 gene generates both a receptor and  
RT its cognate binding protein." ;  
RL Cancer Res. 59:1552-1561(1999).  
RN [21]  
RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.  
RX MEDLINE=21240104; PubMed=11341784;  
RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,  
RA Harris A.;  
RA "Identification of MUC1 proteolytic cleavage sites in vivo." ;  
RL Biochem. Biophys. Res. Commun. 283:715-720(2001).  
RN [22]  
RP CHARACTERIZATION.  
RX MEDLINE=21836452; PubMed=11847293;  
RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,  
RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,  
RA Stacey M., Lin H.-H., Gordon S.;  
RA "Generation of ligand-receptor alliances by 'SEA' module-mediated  
RT cleavage of membrane-associated mucin proteins." ;  
RL Protein Sci. 11:698-706(2002).  
RN [23]  
RP PHOSPHORYLATION.  
RX MEDLINE=95080414; PubMed=7988707;  
RA Zrihan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.;  
RA "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.  
RT Cytokine receptor-like molecules." ;  
RL FEBS Lett. 356:130-136(1994).  
CC -!- FUNCTION: May play a role in adhesive functions and in cell-cell  
CC interactions, metastasis and signaling. May provide a protective

Query Match 100.0%; Score 127; DB 1; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAHGVTSPD 23  
Db 142 PDTRPAGSTAPPAHGVTSPD 164

RESULT 2  
MUC1\_HYLLA  
ID MUC1\_HYLLA STANDARD; PRT; 475 AA.  
AC Q29435;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Mucin 1 precursor (MUC-1).  
GN MUC1.  
OS Hylobates lar (Common gibbon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.  
OX NCBI\_TaxID=9580;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96351712; PubMed=8747930;  
RA Spicer A.P., Duhig T., Chilton B.S., Gendler S.J.;  
RT "Analysis of mammalian MUC1 genes reveals potential functionally  
RT important domains";  
RL Mamm. Genome 6:885-888(1995).  
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
CC -!- CYTOSKELETON (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.  
CC -!- SIMILARITY: Contains 1 SEA domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L41589; AAA69965.1; -  
EMBL; L41625; AAA69918.1; -  
EMBL; L41624; AAA69918.1; JOINED.  
InterPro; IPR000082; SEA domain.  
Pfam; PF01390; SEA; 1.  
SMART; SM00200; SEA; 1.  
PROSITE; PS50024; SEA; 1.  
Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  
Repeat.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 475 MUCIN 1.  
FT DOMAIN 24 380 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 381 401 POTENTIAL.  
FT DOMAIN 402 475 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 102 121 1.  
FT REPEAT 122 141 2.  
FT REPEAT 142 161 3.  
FT REPEAT 162 181 4.  
FT DOMAIN 254 371 SEA.  
SQ SEQUENCE 475 AA; 49371 MW; D7A699D6D68C6622 CRC64;

Query Match 89.0%; Score 113; DB 1; Length 475;  
Best Local Similarity 91.3%; Pred. No. 2.1e-06;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAHGVTSPD 23  
Db 122 PDTRPAGSTAPPAHGVTSPD 144

RESULT 3  
MSL1\_DROME  
ID MSL1\_DROME STANDARD; PRT; 1039 AA.  
AC P50535;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Male-specific lethal-1 protein.  
GN MSL-1.  
OS Drosophila melanogaster (Fruit fly).  
OC Sukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE OF 85-1039 FROM N.A.  
RC STRAIN=Canton-S;  
RX MEDLINE=93314941; PubMed=8325488;  
RA Palmer M.J., Mergner V.A., Richman R., Manning J.E., Kuroda M.I.,  
RA Lucchesi J.C.;  
RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster  
RT encodes a novel protein that associates with the X chromosome in  
RT males.";  
RL Genetics 134:545-557(1993).  
RN [2]  
RP REVISIONS, SEQUENCE FROM N.A.  
RX MEDLINE=95300219; PubMed=7781064;  
RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,  
RA Kuroda M.I.;  
RT "Expression of msl-2 causes assembly of dosage compensation  
RT regulators on the X chromosomes and female lethality in Drosophila.";  
RL Cell 81:867-877(1995).

CC -!- FUNCTION: The msl proteins are essential for elevating  
CC transcription of the single X chromosome in the male (X chromosome  
CC dosage compensation). msl-1 is a pioneer protein. msl-1 and  
CC msl-3 are co-localized on the X chromosome. Each of the msl  
CC proteins requires all the other msls for wild-type X-chromosome  
CC binding.  
CC -!- SUBUNIT: msl-1 seems to form a tight complex with msl-2.  
CC -!- SUBCELLULAR LOCATION: Nuclear; msl-1 is associated with hundreds  
CC of discrete sites along the length of the X chromosome in males  
CC and not in females, and is also associated with 10-20 autosomal  
CC sites in males.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L42514; AAA98918.1; -  
PIR; S52959; S52959.  
FlyBase; FBgn005617; msl-1.  
GO; GO:0016456; C:dosage compensation complex (sensu Drosophila); NAS.  
GO; GO:0003682; F:chromatin binding; IDA.  
GO; GO:0009047; P:dosage compensation, by hyperactivation of . . .; NAS.  
KW Nuclear protein.  
FT CONFLICT 188 193 PLPPAA -> HCHLLP (IN REF. 1).  
FT CONFLICT 492 492 L -> S (IN REF. 1).  
FT CONFLICT 670 670 I -> M (IN REF. 1).  
SQ SEQUENCE 1039 AA; 117412 MW; 4759E9B5EF6E9F14 CRC64;

Query Match 48.0%; Score 61; DB 1; Length 1039;  
Best Local Similarity 59.1%; Pred. No. 6.1;  
Matches 13; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

QY 1 PDTRPAGSTAPPAHGVTSPD 22  
Db 748 PSTTTPGST--PQHAVTSSMD 767

```
RESULT 4
MM09_CANFA          STANDARD;          PRT;   704 AA.
ID  MM09_CANFA
AC  Q18733;
DT  15-JUL-1998 (Rel. 36, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  92 kDa type IV collagenase precursor (EC 3.4.24.35) (92 kDa
DE  gelatinase) (Matrix metalloproteinase-9) (MMP-9) (Gelatinase B)
DE  (GELB).
GN  MMP9.
OS  Canis familiaris (Dog).
OC  Sukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX  NCBI_TaxID=9615;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Mongrel;
RX  MEDLINE=21587576; PubMed=11731079;
RA  Yokota H., Kumata T., Taketaba S., Kobayashi T., Moue H., Taniyama H.,
RA  Hirayama K., Kagawa Y., Itoh N., Fujita O., Nakade T., Yuasa A.;
RT  "High expression of 92 kDa type IV collagenase (matrix
RT  metalloproteinase-9) in canine mammary adenocarcinoma.";
RL  Biochim. Biophys. Acta 1568:7-12(2001).
CC  -!- FUNCTION: Could play a role in bone osteoclastic resorption.
CC  -!- CATALYTIC ACTIVITY: Cleavage of gelatin types I and V and collagen
CC  types IV and V.
CC  -!- COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).
CC  -!- SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II
CC  MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN
CC  BINDS GELATIN.
CC  -!- SIMILARITY: Belongs to peptidase family M10A.
CC  -!- SIMILARITY: Contains 1 hemopexin-like domain.
CC  -!- SIMILARITY: Contains 3 fibronectin type II domains.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AB006421; BAA22087.3; -.
DR  HSP; P08254; I88Y.
DR  MEROPS; M10.004; -.
DR  InterPro; IPR000562; FN_Type_II.
DR  InterPro; IPR000585; Hemopexin.
DR  InterPro; IPR001818; Pept_M10A_M12B.
DR  InterPro; IPR006025; Pept_M_Zn_BS.
DR  InterPro; IPR006026; Peptidase_M.
DR  InterPro; IPR006970; PT.
DR  Pfam; PF00040; fn2; 3.
DR  Pfam; PF00045; hemopexin; 4.
DR  Pfam; PF00413; Peptidase_M10; 1.
DR  Pfam; PF03933; Peptidase_M10_N; 1.
DR  Pfam; PF04886; FT; 1.
DR  PRINTS; PR00013; FNTYPEII.
DR  PRINTS; PR00138; MATRIXIN.
DR  ProDom; PD000995; FN_Type_II; 3.
DR  SMART; SM00059; FN2; 3.
DR  SMART; SM00120; HX; 4.
DR  SMART; SM00235; ZnMc; 1.
DR  PROSITE; PS00142; ZINC_PROTEASE; 1.
DR  PROSITE; PS00546; CYSTEINE_SWITCH; 1.
DR  PROSITE; PS00023; FIBRONECTIN_2; 3.
DR  PROSITE; PS00024; HEMOPEXIN; FALSE NEG.
KW  Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;
KW  Collagen degradation; Extracellular matrix; Repeat; Signal.
FT  SIGNAL          1 19          BY SIMILARITY.
FT  PROPEP          20 106        ACTIVATION PEPTIDE (BY SIMILARITY).
FT  CHAIN           107 704       92 kDa TYPE IV COLLAGENASE.
```

```
FT  DOMAIN          223 280      FIBRONECTIN TYPE-II 1.
FT  DOMAIN          231 339      FIBRONECTIN TYPE-II 2.
FT  DOMAIN          340 397      FIBRONECTIN TYPE-II 3.
FT  DOMAIN          445 508      PRO-RICH.
FT  DOMAIN          510 704      HEMOPEXIN-LIKE.
FT  METAL           401 401      ZINC (CATALYTIC) (BY SIMILARITY).
FT  ACT_SITE        402 402      BY SIMILARITY.
FT  METAL           405 405      ZINC (CATALYTIC) (BY SIMILARITY).
FT  METAL           411 411      ZINC (CATALYTIC) (BY SIMILARITY).
FT  CARBOHYD         38 38        N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD        127 127      N-LINKED (GLCNAC. .) (POTENTIAL).
FT  DISULFID        513 701      BY SIMILARITY.
SQ  SEQUENCE        704 AA; 78123 MW; 0DB394D2D6256B91 CRC64;

Query Match          45.7%; Score 58; DB 1; Length 704;
Best Local Similarity 52.4%; Pred.No. 9.4;
Matches 11; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY  1 PDTRPAPGSTAPPAHGVTSAP 21
Db  474 PSERPATAGTGPAPAGTGP 494

RESULT 5
Y966_TREPA
ID  Y966_TREPA      STANDARD;          PRT;   544 AA.
AC  O83932;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Hypothetical protein TP0966 precursor.
GN  TP0966.
OS  Treponema pallidum.
OC  Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX  NCBI_TaxID=160;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Nichols;
RX  MEDLINE=98332770; PubMed=9665876;
RA  Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA  Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA  Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA  Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA  McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA  Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA  Venter J.C.;
RT  "Complete genome sequence of Treponema pallidum, the syphilis
RT  spirochete.";
RL  Science 281:375-388(1998).
CC  -!- SIMILARITY: BELONGS TO THE TP096X FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE001264; AAC65924.1; -.
DR  PIR; A71260; A71260.
DR  TIGR; TP0966; -.
KW  Hypothetical protein; Signal; Complete proteome.
FT  SIGNAL          1 34          POTENTIAL.
FT  CHAIN           35 544        HYPOTHETICAL PROTEIN TP0966.
SQ  SEQUENCE        544 AA; 60147 MW; 123D5C731F3E6F11 CRC64;

Query Match          44.9%; Score 57; DB 1; Length 544;
Best Local Similarity 62.5%; Pred.No. 9.5;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  5 PAPGSTAPPAHGVTSAP 20
||||: ||||| : |
```

Db 40 PAPGTPAPPAHTASEA 55

RESULT 6

MM09 MOUSE

ID MM09 MOUSE STANDARD; PRT; 730 AA.

AC P41245; Q06788; Q9DC02;

DT 01-FEB-1995 (Rel. 31, last sequence update)

DT 01-FEB-1995 (Rel. 31, last sequence update)

DT 10-OCT-2003 (Rel. 42, last annotation update)

DE 92 kDa type IV collagenase precursor (EC 3.4.24.35) (92 kDa gelatinase) (Matrix metalloproteinase-9) (MMP-9) (Gelatinase B) (GELB).

DE (GELB).

GN MMP9 OR CLG4B.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Bone;

RX MEDLINE=94179406; PubMed=8132709;

RA Reponen P., Sahlberg C., Munaut C., Thesleff I., Tryggvason K.; "High expression of 92-kD type IV collagenase (gelatinase B) in the osteoclast lineage during mouse development.";

RT J. Cell Biol. 124:1091-1102(1994).

RL [2]

RN SEQUENCE FROM N.A.

RX MEDLINE=93176173; PubMed=8382489;

RA Tanaka H., Hojo K., Yoshida H., Yoshioka T., Sugita K.;

RT "Molecular cloning and expression of the mouse 105-kDa gelatinase cDNA.";

RL Biochem. Biophys. Res. Commun. 190:732-740(1993).

RN [3]

RP SEQUENCE FROM N.A.

RX STRAIN=C57BL/6; TISSUE=Liver;

RX MEDLINE=94062823; PubMed=8243459;

RA Masure S., Nys G., Fiten P., van Damme J., Opdenakker G.;

RT "Mouse gelatinase B. cDNA cloning, regulation of expression and glycosylation in WEHI-3 macrophages and gene organisation.";

RL Eur. J. Biochem. 218:129-141(1993).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=94033534; PubMed=8219207;

RA Graubert T., Johnston J., Berliner N.;

RT "Cloning and expression of the cDNA encoding mouse neutrophil gelatinase: demonstration of coordinate secondary granule protein gene expression during terminal neutrophil maturation.";

RL Blood 82:3192-3197(1993).

RN [5]

RP SEQUENCE FROM N.A.

RX STRAIN=C57BL/6J; TISSUE=Lung;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S., Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

CC -!- FUNCTION: Could play a role in bone osteoclastic resorption.

CC -!- CATALYTIC ACTIVITY: Cleavage of gelatin types I and V and collagen types IV and V.

CC -!- COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).

CC -!- SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN BINDS GELATIN.

CC -!- SIMILARITY: Belongs to peptidase family M10A.

CC -!- SIMILARITY: Contains 1 hemopexin-like domain.

CC -!- SIMILARITY: Contains 3 fibronectin type II domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Z27231; CAA81745.1; -.

DR EMBL; D12712; BAA02208.1; -.

DR EMBL; X72794; CAA51314.1; -.

DR EMBL; X72795; CAA51315.1; -.

DR EMBL; S67830; AAB28942.1; -.

DR EMBL; AK004651; BAB23442.1; -.

DR PIR; JCI456; JCI456.

DR HSSP; P08254; IUSN.

DR MEROPS; M10.004; -.

DR MGD; MGI:97011; Mmp9.

DR InterPro; IPR000562; FN\_Type\_II.

DR InterPro; IPR000585; Hemopexin.

DR InterPro; IPR001818; Pept\_M10A\_M12B.

DR InterPro; IPR006025; Pept\_M\_Zn\_BS.

DR InterPro; IPR006026; Peptidase\_M.

DR InterPro; IPR006970; PT.

DR Pfam; PF00040; fn2; 3.

DR Pfam; PF00045; hemopexin; 4.

DR Pfam; PF00413; Peptidase\_M10; 1.

DR Pfam; PF03933; Peptidase\_M10\_N; 1.

DR Pfam; PF04886; PT; 2.

DR PRINTS; PR0013; FNTYPEII.

DR PRINTS; PR00138; MATRININ.

DR ProDom; PD000995; FN\_Type\_II; 3.

DR SMART; SM00059; FN2; 3.

DR SMART; SM00120; HX; 4.

DR SMART; SM00235; ZnMc; 1.

DR PROSITE; PS00023; FIBRONECTIN 2; 3.

DR PROSITE; PS00024; HEMOPEXIN; 1.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

DR PROSITE; PS00546; CYSTEINE\_SWITCH; 1.

KW Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;

KW Collagen degradation; Extracellular matrix; Repeat; Signal;

KW Polymorphism.

FT SIGNAL 1 19 BY SIMILARITY.

FT PROPEP 20 107 ACTIVATION PEPTIDE (BY SIMILARITY).

FT CHAIN 108 730 92 kDa TYPE IV COLLAGENASE.

FT DOMAIN 223 280 FIBRONECTIN TYPE-II 1.

FT DOMAIN 281 339 FIBRONECTIN TYPE-II 2.

FT DOMAIN 340 397 FIBRONECTIN TYPE-II 3.

FT DOMAIN 531 730 HEMOPEXIN-LIKE.

FT SITE 100 100 CYSTEINE SWITCH (BY SIMILARITY).

FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT SITE 402 402 BY SIMILARITY.

FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).

FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DISULFID 534 729 BY SIMILARITY.

FT VARIANT 514 514 A -> P.

FT VARIANT 639 639 P -> L.

FT VARIANT 711 711 P -> H.

FT CONFLICT 20 20 A -> C (IN REF. 4).



```

FT CONFLICT 25 26 QP -> HA (IN REF. 4).
FT CONFLICT 466 466 P -> T (IN REF. 5).
FT CONFLICT 514 514 A -> P (IN REF. 5).
FT CONFLICT 639 639 P -> L (IN REF. 5).
FT CONFLICT 711 711 P -> H (IN REF. 5).
SQ SEQUENCE 730 AA; 80453 MW; E1911F6D5CCAC059 CRC64;

Query Match 44.9%; Score 57; DB 1; Length 730;
Best Local Similarity 47.6%; Pred. No. 13;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAHGVTSP 21
Db 495 PTSSPSPGTGAAPSPGTAAP 515

RESULT 7
NUKM ARATH STANDARD; PRT; 218 AA.
AC Q42577;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE NADH-ubiquinone oxidoreductase 20 kDa subunit, mitochondrial precursor
DE (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-20KD) (CI-20KD).
GN AT5G11770 OR T22P22_160.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. C24;
RX MEDLINE=97071689; PubMed=8914535;
RA Heiser V., Grohmann L., Brennicke A.;
RT "The plant mitochondrial 22 kDa (PSST) subunit of respiratory chain
RT complex I is encoded by a nuclear gene with enhanced transcript
RT levels in flowers."
RL Plant Mol. Biol. 31:1195-1204(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Korani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki K., Naruo K., Okumura S., Shinozaki S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang B., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana."
RL Nature 408:823-826(2000).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.

```

```

CC -!- COFACTOR: Binds 1 4Fe-4S cluster (Potential).
CC -!- SUBUNIT: Complex I is composed of about 40 different subunits.
CC -!- SIMILARITY: Belongs to the complex I 20 kDa subunit family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X84078; CAA58887.1; -.
CC EMBL; AL163814; CAB87695.1; -.
CC PIR; S52286; S52286.
CC InterPro; IPR006138; Cmp1x1_20kDa.
CC InterPro; IPR006137; Oxidored q6.
CC Pfam; PF01058; oxidored q6; 1.
CC PROSITE; PS01150; COMPLEX1_20K; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transit peptide;
KW Iron-sulfur; 4Fe-4S.
FT TRANSIT 1 ? ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 218 NADH-UBIQUINONE OXIDOREDUCTASE 20 kDa
FT SUBUNIT.
FT METAL 93 93 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 94 94 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 158 158 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 188 188 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 218 AA; 24044 MW; D4FC0E15A4029908 CRC64;

Query Match 44.1%; Score 56; DB 1; Length 218;
Best Local Similarity 50.0%; Pred. No. 5;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 TRPAPGSTAPPAHGVTSPD 22
Db 47 TRPGPPSTSPPPGSLKAAE 66

RESULT 8
MM09 RAT STANDARD; PRT; 708 AA.
AC P50282;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 92 kDa type IV collagenase precursor (BC 3.4.24.35) (92 kDa
DE gelatinase) (Matrix metalloproteinase-9) (MMP-9) (Gelatinase B)
DE (GELB).
GN MMP9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer 344;
RX MEDLINE=96184505; PubMed=8605986;
RA Xia Y., Garcia G., Chen S., Wilson C.B., Feng L.;
RT "Cloning of rat 92-kDa type IV collagenase and expression of an
RT active recombinant catalytic domain."
RL FEBS Lett. 382:285-288(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=96069602; PubMed=7590350;
RA Okada A., Santavica M., Basset P.;
RT "The cDNA cloning and expression of the gene encoding rat gelatinase
RT B."
RL Gene 164:317-321(1995).
CC -!- FUNCTION: Could play a role in bone osteoclastic resorption.
CC -!- CATALYTIC ACTIVITY: Cleavage of gelatin types I and V and collagen
CC types IV and V.

```



CC CC COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).  
 CC CC SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II  
 CC CC MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN  
 CC CC BINDS GELATIN.  
 CC CC SIMILARITY: Belongs to peptidase family M10A.  
 CC CC SIMILARITY: Contains 1 hemopexin-like domain.  
 CC CC SIMILARITY: Contains 3 fibronectin type II domains.  
 CC CC  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC CC -----  
 CC CC EMBL; U36476; AAB01721.1; --  
 CC CC EMBL; U24441; AAA90911.1; --  
 CC CC PIR; JC4364; JC4364.  
 CC CC PIR; S62907; S62907.  
 CC CC HSSP; P08254; IUSN.  
 CC CC MEROPS; M10\_004; --  
 CC CC InterPro; IPR000562; FN Type II.  
 CC CC InterPro; IPR000585; Hemopexin.  
 CC CC InterPro; IPR001818; Pept\_M10A\_M12B.  
 CC CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC CC InterPro; IPR006026; Peptidase\_M.  
 CC CC InterPro; IPR006970; PT.  
 CC CC Pfam; PF00040; fn2; 3.  
 CC CC Pfam; PF00045; hemopexin; 4.  
 CC CC Pfam; PF00413; Peptidase\_M10; 1.  
 CC CC Pfam; PF03933; Peptidase\_M10\_N; 1.  
 CC CC Pfam; PF04886; PT; 1.  
 CC CC PRINTS; PR00013; FNTYPEII.  
 CC CC PRODom; PR00138; MATRIXIN.  
 CC CC ProDom; PR000995; FN Type II; 3.  
 CC CC SMART; SM00059; FN2; 3.  
 CC CC SMART; SM00120; HX; 4.  
 CC CC SMART; SM00235; ZnMc; 1.  
 CC CC PROSITE; PS00023; FIBRONECTIN\_2; 2.  
 CC CC PROSITE; PS00024; HEMOPEXIN; 1.  
 CC CC PROSITE; PS00142; ZINC PROTEASE; 1.  
 CC CC PROSITE; PS00546; CYSTEINE SWITCH; 1.  
 CC CC Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;  
 CC CC Collagen degradation; Extracellular matrix; Repeat; Signal.  
 CC CC SIGNAL 1 19 BY SIMILARITY.  
 CC CC PROPEP 20 107 ACTIVATION PEPTIDE (BY SIMILARITY).  
 CC CC CHAIN 108 708 92 kDa TYPE IV COLLAGENASE.  
 CC CC DOMAIN 224 281 FIBRONECTIN TYPE-II 1.  
 CC CC DOMAIN 282 340 FIBRONECTIN TYPE-II 2.  
 CC CC DOMAIN 341 398 FIBRONECTIN TYPE-II 3.  
 CC CC DOMAIN 516 708 HEMOPEXIN-LIKE.  
 CC CC SITE 100 100 CYSTEINE SWITCH (BY SIMILARITY).  
 CC CC METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC CC ACT\_SITE 403 403 BY SIMILARITY.  
 CC CC METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC CC METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC CC CARBOHYD 39 39 N-LINKED GLCNAC. . . (POTENTIAL).  
 CC CC CARBOHYD 121 121 N-LINKED GLCNAC. . . (POTENTIAL).  
 CC CC DISULFID 519 707 BY SIMILARITY.  
 CC CC CONFLICT 2 2 S -> N (IN REF. 2).  
 CC CC CONFLICT 112 112 D -> E (IN REF. 2).  
 CC CC CONFLICT 326 327 AD -> LY (IN REF. 2).  
 CC CC CONFLICT 364 364 S -> Q (IN REF. 2).  
 CC CC CONFLICT 441 441 H -> G (IN REF. 2).  
 CC CC CONFLICT 472 472 S -> P (IN REF. 2).  
 CC CC CONFLICT 515 515 D -> V (IN REF. 2).  
 CC CC CONFLICT 551 551 N -> S (IN REF. 2).  
 CC CC CONFLICT 566 566 F -> L (IN REF. 2).  
 CC CC CONFLICT 568 568 S -> A (IN REF. 2).  
 CC CC CONFLICT 579 579 P -> S (IN REF. 2).  
 CC CC CONFLICT 586 589 LWAQ -> SGRK (IN REF. 2).  
 CC CC CONFLICT 597 597 S -> T (IN REF. 2).

FT CONFLICT 669 669 Q -> H (IN REF. 2).  
 SQ SEQUENCE 708 AA; 78610 MW; D57DC0D1E93A778C CRC64;  
 Query Match 44.1%; Score 56; DB 1; Length 708;  
 Best Local Similarity 47.6%; Pred. No. 16;  
 Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
 Qy 1 PDTRPAPGSGTAPPAGHGVTSAP 21  
 Db 488 PTGAPSGPTGPPTAGESEAP 508  
 RESULT 9  
 MM09\_HUMAN  
 ID MM09\_HUMAN STANDARD; PRT; 707 AA.  
 AC P14780; Q8N725; Q9H4Z1;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE 92 kDa type IV collagenase precursor (EC 3.4.24.35) (92 kDa  
 DE gelatinase) (Matrix metalloproteinase-9) (MMP-9) (Gelatinase B)  
 DE (GELB).  
 GN MMP9 OR CLG4B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Sutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.  
 RX MEDLINE=90008879; PubMed=2551898;  
 RA Wilhelm S.M., Collier I.B., Marmer B.L., Eisen A.Z., Grant G.A.,  
 RA Goldberg G.I.;  
 RT "SV40-transformed human lung fibroblasts secrete a 92-kDa type IV  
 RT collagenase which is identical to that secreted by normal human  
 RT macrophages.";  
 RL J. Biol. Chem. 264:17213-17221(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91358433; PubMed=1653238;  
 RA Huhtala P., Tuuttila A., Chow L.T., Lohi J., Keski-Oja J.,  
 RA Tryggvason K.;  
 RT "Complete structure of the human gene for 92-kDa type IV collagenase.  
 RT Divergent regulation of expression for the 92- and 72-kilodalton  
 RT enzyme genes in HT-1080 cells.";  
 RL J. Biol. Chem. 266:16485-16490(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehaeslahti M.H., Liversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A., AND VARIANT GLN-279.  
 RA Rieder M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,  
 RA Rajkumar N.R., Teth E.J., Yi Q., Nickerson D.A.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=B-cell;  
 RX MEDLINE=2238257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano M.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RP SEQUENCE OF 1-11 FROM N.A.  
 RX MEDLINE=93149601; PubMed=8426746;  
 RA Sato H., Seiki M.;  
 RT "Regulatory mechanism of 92 kDa type IV collagenase gene expression  
 which is associated with invasiveness of tumor cells.";  
 RL Oncogene 8:395-405(1993).  
 RN [7]  
 RP SEQUENCE OF 20-37.  
 RX MEDLINE=91355647; PubMed=1653055;  
 RA van Ranst M., Norga K., Masure S., Proost P., Vandekerckhove F.,  
 RA Auwerx J., van Damme J., Opdenakker G.;  
 RT "The cytokine-protease connection: identification of a 96-kD THP-1  
 gelatinase and regulation by interleukin-1 and cytokine inducers.";  
 RL Cytokine 3:231-239(1991).  
 RN [8]  
 RP SEQUENCE OF 28-60.  
 RC TISSUE=Neutrophils;  
 RX MEDLINE=91249834; PubMed=1645657;  
 RA Masure S., Proost P., van Damme J., Opdenakker G.;  
 RT "Purification and identification of 91-kDa neutrophil gelatinase.  
 Release by the activating peptide interleukin-8.";  
 RL Eur. J. Biochem. 198:391-398(1991).  
 RN [9]  
 RP CHARACTERIZATION.  
 RA Kang K., Lee D.-H.;  
 RT "Purification and characterization of human 92-kDa type IV collagenase  
 (gelatinase B).";  
 RL Exp. Mol. Med. 28:161-165(1996).  
 RN [10]  
 RP 3D-STRUCTURE MODELING.  
 RA Mallena S.C., Sagajkar R.D.;  
 RT "Theoretical model of human type IV collagenase precursor.";  
 RL Submitted (APR-2002) to the PDB data bank.  
 RN [11]  
 RP VARIANTS VAL-20; LYS-82 AND GLN-279.  
 RX MEDLINE=20065865; PubMed=10598806;  
 RA Zhang B., Henney A., Eriksson P., Hamsten A., Watkins H., Ye S.;  
 RT "Genetic variation at the matrix metalloproteinase-9 locus on  
 chromosome 20q12.2-13.1.";  
 RL Hum. Genet. 105:418-423(1999).  
 CC -!- FUNCTION: Could play a role in bone osteoclastic resorption.

CC -!- CATALYTIC ACTIVITY: Cleavage of gelatin types I and V and collagen  
 CC types IV and V.  
 CC -!- COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).  
 CC -!- SUBUNIT: Exists as monomer, disulfide-linked homodimer, and as a  
 CC heterodimer with a 25 kDa protein. Macrophages and transformed  
 CC cell lines produce only the monomeric form.  
 CC -!- TISSUE SPECIFICITY: Produced by normal alveolar macrophages and  
 CC granulocytes.  
 CC -!- SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II  
 CC MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN  
 CC BINDS GELATIN.  
 CC -!- SIMILARITY: Belongs to peptidase family M10A.  
 CC -!- SIMILARITY: Contains 1 hemopexin-like domain.  
 CC -!- SIMILARITY: Contains 3 fibronectin type II domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J05070; AAA51539.1; -.  
 CC EMBL: D10051; BAA20967.1; -.  
 CC EMBL: AL162458; CAC10459.1; -.  
 CC EMBL: AF538844; AAM97934.1; -.  
 CC EMBL: BC006093; AAH06093.1; -.  
 CC PIR: A34458; A34458.  
 CC PDB: 1LKG; 15-MAY-02.  
 CC PDB: 1GKC; 16-MAY-02.  
 CC PDB: 1GKD; 16-MAY-02.  
 CC PDB: 1ITV; 04-SEP-02.  
 CC MEROPS: M10.004; -.  
 CC GlycoSuiteDB: P14780; -.  
 CC Genew: HGNC:7176; MMP9.  
 CC MIM: 120361; -.  
 CC GO: GO:0005615; C:extracellular space; TAS.  
 CC GO: GO:0008133; F:collagenase activity; TAS.  
 CC GO: GO:0008270; F:zinc ion binding; TAS.  
 CC GO: GO:0006503; P:proteolysis and peptidolysis; TAS.  
 CC InterPro: IPR000562; FN\_Type\_II.  
 CC InterPro: IPR000585; Hemopexin.  
 CC InterPro: IPR01818; Pept\_M10A\_M12B.  
 CC InterPro: IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro: IPR006026; Peptidase\_M.  
 CC InterPro: IPR006970; PT.  
 CC Pfam: PF00040; fn2; 3.  
 CC Pfam: PF00045; hemopexin; 4.  
 CC Pfam: PF00413; Peptidase\_M10; 1.  
 CC Pfam: PF03933; Peptidase\_M10\_N; 1.  
 CC Pfam: PF04886; PT; 1.  
 CC PRINTS: PR00013; FNTYPEII.  
 CC PRINTS: PR00138; MATRIXIN.  
 CC ProDom: PD000995; FN\_Type\_II; 3.  
 CC SMART: SM00059; FN2; 3.  
 CC SMART: SM00120; HX; 4.  
 CC SMART: SM00235; ZnMc; 1.  
 CC PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 CC PROSITE: PS00023; FIBRONECTIN\_2; 3.  
 CC PROSITE: PS00024; HEMOPEXIN; 1.  
 CC PROSITE: PS00546; CYSTEINE\_SWITCH; 1.  
 CC KW Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;  
 CC KW Collagen degradation; Extracellular matrix; Repeat; Signal;  
 CC KW Polymorphism; 3D-structure.  
 CC FT SIGNAL 1 19  
 CC FT PROPEP 20 106 ACTIVATION PEPTIDE.  
 CC FT CHAIN 107 707 92 kDa TYPE IV COLLAGENASE.

Query Match 43.3%; Score 55; DB 1; Length 707;  
 Best Local Similarity 47.6%; Pred. No. 21;  
 Matches 10; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTSAP 21  
 Db 477 PSERPTAGTGPSPSAGTGP 497

RESULT 10  
 ZYX\_MOUSE  
 ID ZYX\_MOUSE STANDARD; PRT; 564 AA.  
 AC Q62523; P70461;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Zyxin.  
 GN Zyxin.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR X Swiss Webster;  
 RX MEDLINE=97094926; PubMed=8940160;  
 RA Macalusa T., Otte J., Hensler M.E., Bockholt S.M., Louis H.A.,  
 RA Kalf-Suske M., Grzeschik K.H., von der Ahe D., Beckerle M.C.;  
 RT "Molecular characterization of human zyxin.";  
 RL J. Biol. Chem. 271:31470-31478(1996).  
 RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Otte J., Heischmann A., Breier G., Beckerle M.C., von der Ahe D.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Adhesion plaque protein. Binds alpha-actinin and the CRP

CC protein. May be a component of a signal transduction pathway that

CC mediates adhesion-stimulated changes in gene expression (By

CC similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic; associates with the actin

CC cytoskeleton near the adhesion plaques.

CC -!- SIMILARITY: Contains 3 LIM zinc-binding domains.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Y07711; CAA68984.1; -

DR EMBL; X99063; CAA67510.1; -

DR MGD; MGI:103072; Zyx.

DR InterPro; IPR001781; LIM.

DR Pfam; PF00412; LIM; 3.

DR ProDom; PD000094; LIM; 3.

DR SMART; SM00132; LIM; 3.

DR PROSITE; PS00478; LIM\_DOMAIN\_1; 2.

DR PROSITE; PS50023; LIM\_DOMAIN\_2; 3.

KW Repeat; LIM domain; Metal-binding; Zinc; Cell adhesion.

FT DOMAIN 64 77 PRO-RICH.

FT DOMAIN 94 138 PRO-RICH.

FT DOMAIN 376 435 LIM 1.

FT DOMAIN 436 495 LIM 2.

FT DOMAIN 496 562 LIM 3.

FT CONFLICT 215 215 R -> A (IN REF. 1).

FT CONFLICT 284 292 IKKWCLEMP -> NQKMWPPDA (IN REF. 1).

FT CONFLICT 484 484 S -> C (IN REF. 1).

SQ SEQUENCE 564 AA; 60790 MW; 001E1B3C82ADA1EB CRC64;

Query Match

Best Local Similarity 42.5%; Score 54; DB 1; Length 564;

Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAP 12

| : | | | | |

Db 186 PSTKPAPGCTAP 197

RESULT 11

K685 HUMAN

ID K685 HUMAN STANDARD; PRT; 927 AA.

AC O75170; Q9UGB9;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Protein KIAA0685.

GN KIAA0685.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=98403880; PubMed=9734811;

RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,

RA Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. X.

RT The complete sequences of 100 new cDNA clones from brain which can

RT code for large proteins in vitro.";

RL DNA Res. 5:169-176(1998).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=20057155; PubMed=10591208;

RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,

RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,

RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,

RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,

RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,

RA Clegg S.M., Copley V.E., Cole C.G., Collier R.E., Connor R.,

RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,

RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,

RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,

RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,

RA Hunt R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,

RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,

RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,

RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,

RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,

RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,

RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,

RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,

RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,

RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,

RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,

RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,

RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,

RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,

RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,

RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,

RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,

RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,

RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,

RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,

RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,

RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,

RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlffing T.,

RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,

RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,

RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,

RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,

RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,

RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,

RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,

RA Khan A.S., Lane L., Tilahun Y., Wright H.;

RT "The DNA sequence of human chromosome 22.";

RL Nature 402:489-495(1999).

CC -!- SIMILARITY: Belongs to the SAPS family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB014585; BAA31660.1; -.  
 CC EMBL; AL096767; CAB63062.1; -.  
 CC PIR; T00357; T00357.  
 CC InterPro; IPR007587; SAPS.  
 CC Pfam; PF04499; SAPS; 1.  
 CC DOMAIN 787 792 POLY-SER.  
 CC SEQUENCE 927 AA; 101006 MW; DE1A1A4F2F66A0F9 CRC64;  
 SQ  
 Query Match 42.5%; Score 54; DB 1; Length 927;  
 Best Local Similarity 61.1%; Pred. No. 37;  
 Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 4 RPAPGSTAPPAGHVTSP 21  
 | |||: ||||| ||  
 Db 658 RDAPGAGAPPAGKKKEAP 675.

RESULT 12  
 MGA\_HUMAN STANDARD; PRT; 1856 AA.  
 ID AC Q43451;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Maltase-glucoamylase, intestinal [includes: Maltase (EC 3.2.1.20)  
 DE (Alpha-glucoamylase); Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-  
 DE glucosidase)].  
 GN MGAM OR MGA OR MGAML.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBI TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Small intestine;  
 RX MEDLINE=98112863; PubMed=9446624;  
 RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,  
 RA Sterchi E.E.;  
 RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology  
 RT to sucrose-isomaltase.";  
 RL J. Biol. Chem. 273:3076-3081(1998).  
 RN [2]  
 RP REVISIONS TO 776; 1049; 1100; 1541; 1612 AND 1811.  
 RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,  
 RA Sterchi E.E.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP CHARACTERIZATION.  
 RC TISSUE=Small intestine mucosa;  
 RX MEDLINE=8906802; PubMed=3143729;  
 RA Naim H.Y., Sterchi E.E., Lentze M.J.;  
 RT "Structure, biosynthesis, and glycosylation of human small intestinal  
 RT maltase-glucoamylase.";  
 RL J. Biol. Chem. 263:19709-19717(1988).  
 RN [4]  
 RP SULFATION.  
 RX MEDLINE=88082658; PubMed=3121301;  
 RA Danielson E.M.;  
 RT "Tyrosine sulfation, a post-translational modification of microvillar  
 RT enzymes in the small intestinal enterocyte.";  
 RL EMBO J. 6:2891-2896(1987).  
 CC -!- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION  
 CC WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF  
 CC IMMATURITY OR MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE

DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD  
 MANUFACTURING.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SMALL INTESTINE, GRANULOCYTE, AND  
 CC KIDNEY BUT NOT IN SALIVARY GLAND OR PANCREAS.  
 CC -!- PTM: N- and O-glycosylated.  
 CC -!- PTM: DOES NOT UNDERGO INTRACELLULAR OR EXTRACELLULAR PROTEOLYTIC  
 CC CLEAVAGE.  
 CC -!- PTM: Sulfated (By similarity).  
 CC -!- SIMILARITY: Belongs to family 31 of glycosyl hydrolases.  
 CC -!- SIMILARITY: Contains 2 P-type (trefoil) domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF016833; AAC39568.2; -.  
 CC Genew; HGNC:7043; MGAM.  
 CC MIM; 154360; -.  
 CC GO; GO:0003824; F:catalytic activity; TAS.  
 CC GO; GO:0005983; P:starch catabolism; TAS.  
 CC InterPro; IPR000322; Glyco\_hydro\_31.  
 CC InterPro; IPR000519; P\_trefoil.  
 CC Pfam; PF01055; Glyco\_hydro\_31; 2.  
 CC Pfam; PF00088; trefoil; 2.  
 CC SMART; SM00018; PD; 2.  
 CC PROSITE; PS00129; GLYCOSYL HYDROL\_F31\_1; 2.  
 CC PROSITE; PS00025; P\_TREFOIL; 1.  
 CC Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;  
 CC Glycosidase; Repeat; Signal-anchor; Sulfation.  
 CC INIT MET 0 0  
 CC DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 13 33 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC DOMAIN 34 1856 LUMENAL (POTENTIAL).  
 CC DOMAIN 37 83 SER/THR-RICH.  
 CC DOMAIN 88 132 P-TYPE 1.  
 CC DOMAIN 197 914 MALTASE.  
 CC DOMAIN 952 998 P-TYPE 2.  
 CC DOMAIN 1066 1812 GLUCOAMYLASE.  
 CC ACT\_SITE 1419 1419 BY SIMILARITY.  
 CC DISULFID 89 117 BY SIMILARITY.  
 CC DISULFID 100 116 BY SIMILARITY.  
 CC DISULFID 111 129 BY SIMILARITY.  
 CC DISULFID 965 982 BY SIMILARITY.  
 CC DISULFID 977 995 BY SIMILARITY.  
 CC MOD\_RES 415 415 SULFATION (POTENTIAL).  
 CC MOD\_RES 424 424 SULFATION (POTENTIAL).  
 CC MOD\_RES 1281 1281 SULFATION (POTENTIAL).  
 CC CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 826 826 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 884 884 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 976 976 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 1254 1254 N-LINKED (GLCNAC. . .) (POTENTIAL).



```
FT CARBOHYD 1322 1322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1363 1363 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1387 1387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1602 1602 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1671 1671 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1841 1841 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1846 1846 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1856 AA; 209721 MW; 81E7AA0CABBD07D CRC64;

Query Match 42.5%; Score 54; DB 1; Length 1856;
Best Local Similarity 50.0%; Pred. No. 73;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 3 TRPAPGSTAPPAGHGVTSAPD 22
   |||||
Db 41 TAPDPTGTGTPDPGTTGTPD 60
   |||||

RESULT 13
SYJ1_HUMAN
ID SYJ1_HUMAN STANDARD; PRT; 1575 AA.
AC O43426; O43425;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Synaptotagmin 1 (EC 3.1.3.36) (Synaptic inositol-1,4,5-trisphosphate 5-
DE phosphatase 1).
GN SYJ1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Cerebellum;
EX MEDLINE=98088905; PubMed=9428629;
RA Raffner C., Takei K., Chen H., Ringstad N., Hudson A., Butler M.H.,
RA Salcini A.E., Di Fiore P.P., De Camilli P.;
RT "Synaptotagmin 1: localization on coated endocytic intermediates in
RT nerve terminals and interaction of its 170 kDa isoform with Eps15.";
RL FEBS Lett. 419:175-180(1997).
RN [2]
RP INTERACTION WITH AMPH; SH3GL1; SH3GL2 AND SH3GL3.
EX MEDLINE=20011396; PubMed=10542231;
RA Cestra G., Castagnoli L., Dente L., Minenkova O., Petrelli A.,
RA Migone N., Hoffmueller U., Schneider-Mergener J., Cesareni G.;
RT "The SH3 domains of endophilin and amphiphysin bind to the
RT proline-rich region of synaptotagmin 1 at distinct sites that display
RT an unconventional binding specificity.";
RL J. Biol. Chem. 274:32001-32007(1999).
CC -!- FUNCTION: Inositol 5-phosphatase which has a role in clathrin-
CC mediated endocytosis.
CC -!- CATALYTIC ACTIVITY: 1-phosphatidy1-1D-myo-inositol 4,5-
CC bisphosphate + H(2)O = 1-phosphatidyl-1D-myo-inositol 4-phosphate
CC + phosphate.
CC -!- SUBUNIT: Binds AMPH, SH3GL1, SH3GL2 and SH3GL3.
CC -!- SUBCELLULAR LOCATION: Localized mainly in the cytoplasm (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Synaptotagmin-170;
CC IsoId=O43426-1; Sequence=Displayed;
CC Name=2; Synonyms=Synaptotagmin-145;
CC IsoId=O43426-2; Sequence=VSP 002682, VSP 002683;
CC -!- TISSUE SPECIFICITY: Concentrated at clathrin-coated endocytic
CC intermediates in nerve terminals. Isoform 1 is more enriched than
CC isoform 2 in developing brain as well as non-neuronal cells.
CC -!- DOMAIN: Binds to EPS15 (a clathrin coat-associated protein) via a
CC C-terminal domain containing three Asn-Pro-Phe (NPF) repeats (By
CC similarity).
CC -!- DOMAIN: The C-terminal proline-rich region mediates binding to a
CC variety of SH3 domain-containing proteins including AMPH, SH3GL1,
```

```
CC SH3GL2, SH3GL3 and GRB2.
CC -!- SIMILARITY: In the central section; belongs to the inositol-1,4,5-
CC trisphosphate 5-phosphatase family.
CC -!- SIMILARITY: Contains 1 SAC domain.
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF009040; AAC51922.1; -.
CC EMBL; AF009039; AAC51921.1; -.
CC Genes; HGNC:11503; SYNJ1.
CC MIM; 604297; -.
CC GO; GO:0005803; C:secretory vesicle; TAS.
CC GO; GO:0004445; F:inositol-polyphosphate 5-phosphatase activity; TAS.
CC GO; GO:0006899; P:nonselective vesicle transport; TAS.
CC GO; GO:0006796; P:phosphate metabolism; TAS.
CC GO; GO:0008099; P:synaptic vesicle endocytosis; TAS.
CC InterPro; IPR005135; Exo_endo_phos.
CC InterPro; IPR003000; IPPC.
CC InterPro; IPR000504; RNA_rec_mot.
CC InterPro; IPR002013; Syja_N.
CC Pfam; PF03372; Exo_endo_phos; 1.
CC Pfam; PF02383; Syja_N; 1.
CC SMART; SM00128; IPPC; 1.
CC PROSITE; PS50102; RRM; 1.
CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
CC PROSITE; PS50275; SAC; 1.
CC Hydrolase; Alternative splicing; Repeat; Endocytosis; RNA-binding;
KW Multigene family.
KW DOMAIN 119 442
FT DOMAIN 500 899 SAC.
FT DOMAIN 902 971 CATALYTIC (POTENTIAL).
FT DOMAIN 900 1575 RNA-BINDING (RRM).
FT DOMAIN 1033 1036 PRO-RICH.
FT DOMAIN 1108 1113 POLY-SER.
FT DOMAIN 1126 1129 POLY-PRO.
FT DOMAIN 1487 1490 POLY-GLU.
FT DOMAIN 1540 1546 POLY-PRO.
FT DOMAIN 1396 1419 3 X 3 AA REPEATS OF N-P-F.
FT REPEAT 1396 1398 1.
FT REPEAT 1406 1408 2.
FT REPEAT 1417 1419 3.
FT VARSPPLIC 1306 1311 VKTNGI -> QEOPSG (in isoform 2).
FT VARSPPLIC 1312 1575 /FTId=VSP 002682.
FT VARSPPLIC 1575 1575 Missing (in isoform 2).
SQ SEQUENCE 1575 AA; 173345 MW; 50646F6CC043B9E7 CRC64;

Query Match 41.7%; Score 53; DB 1; Length 1575;
Best Local Similarity 47.8%; Pred. No. 81;
Matches 11; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 PDRPAPGSTAPPAGHGVTSAPDT 23
   |||||
Db 1116 PDRPAPQRPQPPPPSGARSAPT 1138
   |||||

RESULT 14
BCB1_ARATH
ID BCB1_ARATH STANDARD; PRT; 196 AA.
AC Q07488; O82664;
DT 01-FEB-1995 (Rel. 31, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Blue copper protein precursor (Blue copper-binding protein) (AtCBP)
DE (Stellacyanin) (Phytocyanin 1).
GN BCE OR AWI 32 CR AT5G20230 OR P5024.120.
```



OS Arabidopsis thaliana (Mouse-ear cress).  
OC Spermatophyta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Eukaryota; Magnoliophyta; Eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia, and cv. Columbia K85;  
RX MEDLINE=94124044; PubMed=8294044;  
RA van Gysel A., van Montagu M., Inze D.;  
RT "A negatively light-regulated gene from Arabidopsis thaliana encodes  
RT a protein showing high similarity to blue copper-binding proteins.";  
RL Gene 136:79-85(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Massilewskij;  
RA Yang K.Y., Kim C.S., Cho B.H.;  
RT "Characterization of a wound-inducible Arabidopsis gene encoding a  
RT protein homologous to blue copper binding proteins.";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Landsberg erecta;  
RX MEDLINE=20233824; PubMed=10769227;  
RA Honma T., Goto K.;  
RT "The Arabidopsis floral homeotic gene PISTILLATA is regulated by  
RT discrete cis-elements responsive to induction and maintenance  
RT signals.";  
RL Development 127:2021-2030(2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=21016721; PubMed=11130714;  
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,  
RA Nakazaki N., Naruo K., Okumura S., Shirpo S., Takeuchi C., Wada T.,  
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,  
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,  
RA Du H., Edwards J., Fryman J., Haakeness B., Lamar E., Latreille P.,  
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,  
RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,  
RA Parrell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,  
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,  
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,  
RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,  
RA van Staveren M., Dirkse W., Mooijman P., Klein lankhorst R.,  
RA Weitzenegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,  
RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,  
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.;  
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
RT thaliana.";  
RL Nature 408:823-826(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=22954850; PubMed=14593172;  
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
RA Hayashizaki Y., Johnson-Hopson C., Heuan V.W., Iida K., Karnes M.,  
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,

RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
RA Yuan S., Shirozaki K., Davis R.W., Theologis A., Ecker J.R.;  
RT "Empirical analysis of transcriptional activity in the Arabidopsis  
RT genome.";  
RL Science 302:842-846(2003).  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.A.;  
RT "Full-length cDNA from Arabidopsis thaliana.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP GPI-ANCHOR.  
RX MEDLINE=22690167; PubMed=12805588;  
RA Borner G.H., Lilley K.S., Stevens T.J., Dupree P.;  
RT "Identification of glycosylphosphatidylinositol-anchored proteins in  
RT Arabidopsis. A proteomic and genomic analysis.";  
RL Plant Physiol. 132:568-577(2003).  
CC -!- FUNCTION: Probably acts as an electron carrier.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -!- DEVELOPMENTAL STAGE: Maximum levels are found in 35 day old  
CC plantlets when the rosette is mature, consisting of 8-10 fully  
CC expanded leaves, and as the floral stem starts to form. This level  
CC remains constant during the further life span of the plant.  
CC -!- INDUCTION: By dark adaptation. This gives a 20-fold increase in  
CC expression.  
CC -!- SIMILARITY: Contains 1 plastocyanin-like domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; Z15058; CAA78771.1; -;  
DR EMBL; Y18227; CAA77089.1; -;  
DR EMBL; AB035137; BAA86999.1; -;  
DR EMBL; AF296825; -; NOT ANNOTATED\_CDS.  
DR EMBL; AY052681; AAK96585.1; -;  
DR EMBL; AY034986; AAK59491.1; -;  
DR EMBL; AY142577; AAN13146.1; -;  
DR EMBL; AY088549; AAM66081.1; -;  
DR PIR; I39698; I39698.  
DR PIR; T51838; T51838.  
DR HSSP; P29602; 1JER.  
DR InterPro; IPR000923; BlueCu 1.  
DR InterPro; IPR008972; Cupredoxin.  
DR InterPro; IPR003245; Pcyenin like.  
DR Pfam; PF02298; Cu\_bind\_like; 1.  
DR ProDom; PD003122; Pcyenin like; 1.  
DR PROSITE; PS00196; COPPER\_BLUE; 1.  
DR Electron transport; Membrane; Metal-binding; Copper; Signal;  
KW Glycoprotein; GPI-anchor; Lipoprotein.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 174 BLUE COPPER PROTEIN.  
FT PROPEP 175 196 REMOVED IN MATURE FORM (PROBABLE).  
FT DOMAIN 23 118 PLASTOCYANIN-LIKE.  
FT METAL 66 66 COPPER (BY SIMILARITY).  
FT METAL 107 107 COPPER (BY SIMILARITY).  
FT METAL 112 112 COPPER (BY SIMILARITY).  
FT METAL 117 117 COPPER (BY SIMILARITY).  
FT DISULFID 79 113 BY SIMILARITY.  
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT LIPID 174 174 GPI-anchor amidated asparagine  
FT (Probable).  
FT CONFLICT 44 44 T -> S (IN REF. 1).  
FT CONFLICT 134 134 P -> L (IN REF. 1).  
FT CONFLICT 142 142 P -> L (IN REF. 1).  
FT SEQUENCE 196 AA; 20053 MW; 05100B50518F0A56 CRC64;



CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-  
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE  
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER  
CC MEMBRANE PROTEINS (BY SIMILARITY).  
CC  
CC -!- SUBUNIT: Homodimer. Forms a complex with the accessory proteins  
CC exbB and exbD (By similarity).  
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC  
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE  
CC PERIPLASM.  
CC  
CC -!- SIMILARITY: Belongs to the tonB family.  
CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X70139; CAA49716.1; -.  
CC PIR; S28444; S28444.  
CC InterPro; IPR006260; TonB\_C.  
CC TIGRFAMs; TIGR01352; tonB\_Cterm; 1.  
CC Transport; Protein transport; Inner membrane; Periplasmic;  
CC Transmembrane; Signal-anchor; Repeat.  
CC DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 13 32 SIGNAL-ANCHOR (POTENTIAL).  
CC DOMAIN 33 243 PERIPLASMIC (POTENTIAL).  
CC DOMAIN 50 61 6 X 2 AA TANDEM REPEATS OF X-P.  
CC SEQUENCE 243 AA; 25993 MW; 6575C15A147A2774 CRC64;  
CC

DR EMBL; X70139; CAA49716.1; -.  
DR PIR; S28444; S28444.  
DR InterPro; IPR006260; TonB\_C.  
DR TIGRFAMs; TIGR01352; tonB\_Cterm; 1.  
KW Transport; Protein transport; Inner membrane; Periplasmic;  
KW Transmembrane; Signal-anchor; Repeat.  
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 13 32 SIGNAL-ANCHOR (POTENTIAL).  
FT DOMAIN 33 243 PERIPLASMIC (POTENTIAL).  
FT DOMAIN 50 61 6 X 2 AA TANDEM REPEATS OF X-P.  
SQ SEQUENCE 243 AA; 25993 MW; 6575C15A147A2774 CRC64;  
SQ

Query Match 40.2%; Score 51; DB 1; Length 243;  
Best Local Similarity 42.9%; Pred. No. 22;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;  
QY 1 PDTRPAPGSTAPPAHGVTSP 21  
Db 104 PEKKPEPPKEAPTEEVVDAP 124

RESULT 17  
SFPO\_HUMAN  
ID SFPO\_HUMAN STANDARD; PRT; 707 AA.  
AC P23246; P30808;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-  
DE binding protein-associated splicing factor) (PTB-associated splicing  
DE factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).  
GN SFPO OR PSF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.  
RC TISSUE=Fetal brain;  
RX MEDLINE=93194059; PubMed=8449401;  
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;  
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing  
RT factor";  
RL Genes Dev. 7:393-406(1993).  
RN [2]  
RP SEQUENCE OF 312-707 FROM N.A.  
RC TISSUE=Fetal skeletal muscle;  
RX MEDLINE=90091812; PubMed=2480877;  
RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;  
RT "Cloning and characterization of a myoblast cell surface antigen  
RT defined by 24.1D5 monoclonal antibody";  
RL Development 105:723-731(1989).  
RN [3]  
RP SEQUENCE OF 48-68 AND 213-246.

RX MEDLINE=93176127; PubMed=8439294;  
RA Zhang W.-M., Zhang L.-X., Busch R.K., Farres J., Busch H.;  
RT "Purification and characterization of a DNA-binding heterodimer of 52  
RT and 100 kDa from HeLa cells.";  
RL Biochem. J. 290:267-272(1993).  
CC -!- FUNCTION: Essential pre-mRNA splicing factor required early in  
CC spliceosome formation. Binds to the mammalian polypyrimidine  
CC tracts. Forms a complex with the polypyrimidine tract-binding  
CC protein (PTB). Seems to also bind DNA.  
CC -!- SUBUNIT: Heterotetramer of two 52 kDa and two 100 kDa subunits.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=Long;  
CC IsoId=P23246-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=P23246-2; Sequence=VSP\_005855;  
CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.  
CC -!- CAUTION: Was originally (Ref.2) thought to be myoblast cell  
CC surface antigen 24.1D5 and a possible membrane-bound protein  
CC ectokinase.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X70944; CAA50283.1; -.  
CC EMBL; X16850; CAA34747.1; -.  
CC PIR; A46302; A46302.  
CC PIR; S29770; S29770.  
CC HSP; P11940; 1CVJ.  
CC SWISS-2DPAGE; P23246; HUMAN.  
CC Genew; HGNC:10774; SFPQ.  
CC GK; P23246; -.  
CC MIM; 605199; -.  
CC GO; GO:0008243; F:pre-mRNA splicing factor activity; TAS.  
CC GO; GO:0006397; P:mRNA processing; TAS.  
CC GO; GO:0006371; P:mRNA splicing; TAS.  
CC InterPro; IPR000504; RNA\_rec\_mot.  
CC Pfam; PF00076; rrm; 2.  
CC SMART; SM00360; RRM; 2.  
CC PROSITE; PS0102; RRM; 2.  
CC PROSITE; PS0030; RRM\_RNP\_1; 1.  
KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;  
KW Alternative splicing.

FT DOMAIN 297 369 RNA-BINDING (RRM) 1.  
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.  
FT DOMAIN 9 27 3 X 3 AA REPEATS OF R-G-G.  
FT REPEAT 9 11 1.  
FT REPEAT 19 21 2.  
FT REPEAT 25 27 3.  
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.  
FT DOMAIN 10 15 POLY-GLY.  
FT DOMAIN 20 27 POLY-GLY.  
FT DOMAIN 56 65 POLY-PRO.  
FT DOMAIN 67 71 POLY-GLN.  
FT DOMAIN 95 98 POLY-GLN.  
FT DOMAIN 99 103 POLY-PRO.  
FT DOMAIN 184 188 POLY-PRO.  
FT DOMAIN 571 574 POLY-ARG.  
FT DOMAIN 613 616 POLY-GLY.  
FT DOMAIN 635 641 POLY-GLY.  
FT VARSPLIC 663 707 RTERFGQGGAGVGGQGRGMGPAGYGRGREYEGPNK  
FT KPRF -> VRMIDVG (in isoform Short).  
FT /FTId=VSP\_005855.  
FT G -> R (IN REF. 3).  
FT CONFLICT 243 243  
FT SEQUENCE 707 AA; 76149 MW; 608D5EA95E235847 CRC64;  
SQ

```
Query Match 40.2%; Score 51; DB 1; Length 707;
Best Local Similarity 57.1%; Pred. No. 63;
Matches 12; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

QY 1 PDTRPAGSTAPPAGHGVTSAP 21
Db 143 PGSGPGTPTPPPA--VTSAP 161

RESULT 18
MM09 BOVIN STANDARD; PRT; 712 AA.
ID MM09 BOVIN STANDARD; PRT; 712 AA.
AC P52176;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 92 kDa type IV collagenase precursor (EC 3.4.24.35) (92 kDa
DE gelatinase) (Matrix metalloproteinase-9) (MMP-9) (Gelatinase B)
DE (GELB).
GN MMP9.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Leukocyte;
RC MEDLINE=95287902; PubMed=7770085;
RA Baylis H.A., Megson A., Hall R.;
RT "Infection with Theileria annulata induces expression of matrix
RT metalloproteinase 9 and transcription factor AP-1 in bovine
RT leucocytes."
RL Mol. Biochem. Parasitol. 69:211-222(1995).
CC -!- FUNCTION: Could play a role in bone osteoclastic resorption.
CC -!- CATALYTIC ACTIVITY: Cleavage of gelatin types I and V and collagen
CC types IV and V.
CC -!- COFACTOR: Binds 2 zinc ions per subunit, calcium (By similarity).
CC -!- SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II
CC MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN
CC BINDS GELATIN.
CC -!- SIMILARITY: Belongs to peptidase family M10A.
CC -!- SIMILARITY: Contains 1 hemopexin-like domain.
CC -!- SIMILARITY: Contains 3 fibronectin type II domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X78324; CAA55127.1; --
DR PIR; I46031; I46031.
DR HSSP; P22894; IJAP.
DR MEROPS; M10.004; --
DR InterPro; IPR000562; FN Type II.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001818; Pept M10A_M12B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR006026; Peptidase_M.
DR InterPro; IPR006970; PT.
DR Pfam; PF00040; fn2; 3.
DR Pfam; PF00045; hemopexin; 4.
DR Pfam; PF00413; Peptidase M10; 1.
DR Pfam; PF03933; Peptidase_M10_N; 1.
DR Pfam; PF04866; PT; 1.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00138; MATRIXIN.
DR ProDom; PD000995; FN Type II; 3.
DR SMART; SM00059; FN2; 3.
DR SMART; SM00120; HX; 4.

SMART; SM00235; ZnMC; 1.
PROSITE; PS00023; FIBRONECTIN_2; 3.
PROSITE; PS00024; HEMOPEXIN; FALSE_NEG.
PROSITE; PS00142; ZINC PROTEASE; 1.
PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;
KW Collagen degradation; Extracellular matrix; Repeat; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT PROPEP 20 106 ACTIVATION PEPTIDE.
FT CHAIN 107 712 92 kDa TYPE IV COLLAGENASE.
FT DOMAIN 223 280 FIBRONECTIN TYPE-II 1.
FT DOMAIN 281 339 FIBRONECTIN TYPE-II 2.
FT DOMAIN 340 397 FIBRONECTIN TYPE-II 3.
FT DOMAIN 518 712 HEMOPEXIN-LIKE.
FT SITE 99 99 CYSTEINE SWITCH (BY SIMILARITY).
FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 521 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 712 AA; 79087 MW; E7FDA28372AEEOCE CRC64;

Query Match 40.2%; Score 51; DB 1; Length 712;
Best Local Similarity 47.6%; Pred. No. 63;
Matches 10; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAP 21
Db 482 PSEGPITGPTGPPAAGTGP 502

RESULT 19
GSPB_ERWCH STANDARD; PRT; 220 AA.
AC Q01563;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE General secretion pathway protein B.
GN OUTB.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3937;
RX MEDLINE=93086427; PubMed=1453958;
RA Condemine G., Dorel C., Hugouvieux-Cotte-Pattat N., Robert-Baudouy J.;
RT "Some of the out genes involved in the secretion of pectate lyases in
RT Erwinia chrysanthemi are regulated by kdgr.";
RL Mol. Microbiol. 6:3199-3211(1992).
RN [2]
RP REVISIONS.
RA Condemine G.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CUT PROTEINS ARE REQUIRED FOR THE TRANSLOCATION OF
CC PECTATE LYASES AND CELLULASES ACROSS THE OUTER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE EXEB/OUTB/PULB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```



```

DR EMBL; X65265; CAA46373.1; -.
KW Transport; Transmembrane; Inner membrane.
FT TRANSMEM 17 37 POTENTIAL.
SQ SEQUENCE 220 AA; 23486 MW; D094FE2DC0931978 CRC64;

Query Match          39.8%; Score 50.5; DB 1; Length 220;
Best Local Similarity 42.9%; Pred. No. 22;
Matches 15; Conservative 1; Mismatches 6; Indels 13; Gaps 2;

QY 1 PDTRPAP-----GSTAPP-----AHGVTSA PD 22
   |||:|||||
Db 108 PDKPAKLVTGWQTAKEGELPYIAFSAHVYTSAPD 142

RESULT 20
VGLG_BRSVW
ID VGLG_BRSVW STANDARD; PRT; 263 AA.
AC O10687;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Major surface glycoprotein G (Attachment glycoprotein G).
GN G.
OS Bovine respiratory syncytial virus (strain Wbh) (BRS).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
OX NCBI_TaxID=82825;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97288324; PubMed=9143302;
RX Furze J., Roberts S., Wertz G., Taylor G.;
RT "Antigenically distinct G glycoproteins of BRSV strains share a high
RT degree of genetic homogeneity.";
RL Virology 231:48-58 (1997).
CC -!- FUNCTION: Unlike the other paramyxovirus attachment proteins, the
CC respiratory syncytial virus G protein lacks both neuraminidase and
CC hemagglutinating activities.
CC -!- SUBCELLULAR LOCATION: Expressed on the surface of the infected
CC cells and incorporated in the membrane of the virions.
CC -!- SIMILARITY: TO THE G PROTEINS OF EITHER THE SUBGROUP A OR B
CC HRS VIRUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y08717; CAA69967.1; -.
CC HSSP; P22261; 1BRV.
CC InterPro; IPR000925; Glycoprot_G.
CC Pfam; PF00802; Glycoprotein_G; 1.
CC TRANSMEMBRANE; Glycoprotein.
CC DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 38 66 POTENTIAL.
CC DOMAIN 67 263 EXTRACELLULAR (POTENTIAL).
CC DISULFID 173 186 BY SIMILARITY.
CC DISULFID 176 182 BY SIMILARITY.
CC CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 263 AA; 29050 MW; QD06AF7FCB46B858 CRC64;

Query Match          39.8%; Score 50.5; DB 1; Length 263;
Best Local Similarity 39.1%; Pred. No. 27;
Matches 9; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

QY 1 PDTRPAP-GSTAPPAGVTSAPD 22
   |||:|||||
Db 225 PEAKPQPKNTAAPQGGILSSPE 247

```

```

RESULT 21
NUSB_DEIRA
ID NUSB_DEIRA STANDARD; PRT; 192 AA.
AC Q9RSQ7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N utilization substance protein B homolog (NusB protein).
GN NUSB OR DR2067.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.B., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577 (1999).
CC -!- FUNCTION: Involved in the transcription termination process (By
CC similarity).
CC -!- SIMILARITY: Belongs to the nusB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002043; AAF11617.1; -.
CC PIR; E75318; E75318.
CC HSSP; P95020; 1EYV.
CC TIGR; DR2067; -.
CC HAMAP; MF_00073; -; 1.
CC InterPro; IPR006027; NusB.
CC Pfam; PF01029; NusB; 1.
CC TRANSCRIPTION TERMINATION; Complete proteome.
CC SEQUENCE 192 AA; 21147 MW; 8401B62C0B31B4C7 CRC64;

Query Match          39.4%; Score 50; DB 1; Length 192;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 12; Conservative 0; Mismatches 8; Indels 4; Gaps 1;

QY 3 TRPAPGSTAP-----PAHGVTSA PD 22
   |||:|||||
Db 164 TAPAPADTAPADTAPADPAASEPD 187

RESULT 22
ASH2_HUMAN
ID ASH2_HUMAN STANDARD; PRT; 628 AA.
AC Q9UBU3; O60659; O60660; Q96B62;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ASH2-like protein.
GN ASH2L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

```



[1]  
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
TISSUE=Fetal brain;  
MEDLINE=21359618; PubMed=11466562;  
Wang J., Zhou Y., Yin B., Du G., Huang X., Li G., Shen Y., Yuan J.,  
Qiang B.;  
"ASH2L: alternative splicing and downregulation during induced  
megakaryocytic differentiation of multipotential leukemia cell  
lines.";  
J. Mol. Med. 79:399-405(2001).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 1).  
MEDLINE=99321701; PubMed=10393421;  
Ikegawa S., Isonura M., Koshizuka Y., Nakamura Y.;  
"Cloning and characterization of ASH2L and ash2l, human and mouse  
homologs of the Drosophila ash2 gene.";  
Cytogenet. Cell Genet. 84:167-172(1999).  
[3]  
SEQUENCE FROM N.A. (ISOFORM 3).  
TISSUE=Skin;  
MEDLINE=22388257; PubMed=12477932;  
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
-!- FUNCTION: May function as a transcriptional regulator. May play a  
role in hematopoiesis.  
-!- SUBCELLULAR LOCATION: Nuclear (Probable).  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=3;  
Name=1; Synonyms=ASH2L1;  
IsoId=Q9UBL3-1; Sequence=Displayed;  
Name=2; Synonyms=ASH2L2;  
IsoId=Q9UBL3-2; Sequence=VSP\_007577, VSP\_007578;  
Name=3;  
IsoId=Q9UBL3-3; Sequence=VSP\_007577;  
Note=No experimental confirmation available;  
-!- TISSUE SPECIFICITY: Ubiquitously expressed. Predominantly  
expressed in adult heart and testis and fetal lung and liver, with  
barely detectable expression in adult lung, liver, kidney,  
prostate, and peripheral leukocytes.  
-!- SIMILARITY: Contains 1 C4-type zinc finger.  
-!- SIMILARITY: Contains 1 SPRY domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF056718; AAC13564.1; -  
EMBL; AF056717; AAC13563.1; -  
EMBL; AB022785; BAA74520.1; -  
EMBL; AB020982; BAA35127.1; -  
EMBL; BC015936; AAI15936.1; -

DR Genew; HGNC:744; ASH2L.  
DR MIM; 604782; -  
DR GO; GO:0005634; C:nucleus; TAS.  
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti...; TAS.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.  
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.  
DR InterPro; IPR003877; SPRY\_receptor.  
DR Pfam; PF00622; SPRY; 2.  
DR SMART; SMC0449; SPRY; 1.  
KW Alternative splicing; Nuclear protein; Transcription regulation;  
KW Metal-binding; Zinc-finger.  
FT ZN\_FING 117 150 C4-TYPE.  
FT DOMAIN 418 582 SPRY.  
FT VARSPLIC 1 94 Missing (in isoform 2 and isoform 3).  
FT VARSPLIC 541 573 /FTID=VSP\_007577.  
FT VARSPLIC 212 212 Missing (in isoform 2).  
FT CONFLICT 212 212 /FTID=VSP\_007578.  
FT CONFLICT 217 217 Q -> H (IN REF. 1; AAC13564).  
FT CONFLICT 292 292 M -> T (IN REF. 1; AAC13564).  
FT CONFLICT 292 292 S -> T (IN REF. 1; AAC13563).  
FT CONFLICT 351 351 H -> Q (IN REF. 1; AAC13564).  
FT CONFLICT 360 360 L -> I (IN REF. 1; AAC13564).  
FT CONFLICT 369 369 P -> S (IN REF. 1; AAC13564).  
SQ SEQUENCE 628 AA; 68722 MW; 8F5F007430D4B863 CRC64;  
Query Match 39.4%; Score 50; DB 1; Length 628;  
Best Local Similarity 50.0%; Pred. No. 73;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 6 APGSTAPPAGHGVTSAP 21  
Db 36 AAGAAAPPGEIGISAAP 51  
RESULT 23  
KITH\_HSVBM  
ID KITH\_HSVBM STANDARD; PRT; 686 AA.  
AC P33802;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Thymidine kinase (EC 2.7.1.21).  
GN TK OR HXLFI.  
OS Bovine herpes virus type 3 (strain WC11) (Malignant catarrhal fever  
virus).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae.  
OX NCBI\_TaxID=33705;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90351277; PubMed=2167059;  
RA Hsu D., Shih L.M., Zee Y.C.;  
RT "Nucleotide sequence of a 3.5 kilobase fragment of malignant  
catarrhal fever virus strain WC11.";  
RL Arch. Virol. 113:53-60(1990).  
CC -!- CATALYTIC ACTIVITY: ATP + thymidine = ADP + thymidine 5'-  
phosphate.  
CC -!- SIMILARITY: Belongs to the herpesviruses thymidine kinase family.  
DR InterPro; IPR001889; TK herpes.  
DR Pfam; PF00693; TK herpes; 1.  
DR ProDom; PD001519; TK herpes; 1.  
KW Transferase; Kinase; DNA synthesis; ATP-binding.  
FT NP\_BIND 243 250 ATP (PROBABLE).  
SQ SEQUENCE 686 AA; 77073 MW; 06A817D080826A5F CRC64;  
Query Match 39.4%; Score 50; DB 1; Length 686;  
Best Local Similarity 57.1%; Pred. No. 80;  
Matches 12; Conservative 1; Mismatches 6; Indels 2; Gaps 1;  
QY 3 TRPAP--GSTAPPAGHGVTSAP 21  
Db 585 TTPAPPRPATTPRRGVTSAP 605

```
RESULT 24
IF2_RHIME
ID IF2_RHIME STANDARD; PRT; 889 AA.
AC Q92SW4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR R00239 OR SMC02914.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe P., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the IF-2 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AL591783; CAC41676.1; -.
DR HAMAP; MF_00100; -.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR004161; EFTU D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR006847; IF2 N.
DR InterPro; IPR005225; Small GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 2.
DR Pfam; PF04760; IF2_N; 1.
DR PRINTS; PR00929; ATHOOK.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 390 538 G-DOMAIN.
FT NP_BIND 396 403 GTP (BY SIMILARITY).
FT NP_BIND 442 446 GTP (BY SIMILARITY).
FT NP_BIND 496 499 GTP (BY SIMILARITY).
SQ SEQUENCE 889 AA; 96655 MW; 8A213FAED0BA34B5 CRC64;

Query Match 39.4%; Score 50; DB 1; Length 889;
Best Local Similarity 52.4%; Pred. No. 1e+02;
Matches 11; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 2 DTRPAPGSTAPPAHGVTSPD 22
:||||| ||| ||| :|||
```

```
Db 194 ETRPQGRAAPAA--TPAAPD 212

RESULT 25
P121 HUMAN
ID P121 HUMAN STANDARD; PRT; 1229 AA.
AC Q9Y2N3; O75115; Q9Y4S7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE of 121 kDa) (P145).
GN POM121 OR NUP121 OR KIAA0618.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Cordes M., Bauer C., Holmes A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 243-1229 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [3]
SEQUENCE OF 1130-1229 FROM N.A.
RC TISSUE=Uterus;
RA Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Essential component of the nuclear pore complex. The
CC repeat-containing domain may be involved in anchoring components
CC of the pore complex to the pore membrane (By similarity).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC MEMBRANE (BY SIMILARITY).
CC -!- DOMAIN: Contains F-X-F-G repeats.
CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AC006014; AAD28064.1; -.
DR EMBL; AB014518; BAA31593.1; -.
DR EMBL; AL080109; CAB45713.1; -.
DR PIR; T12551; T12551.
DR Genew; HGNC:19702; POM121.
KW Nuclear protein; Transport; Transmembrane; Repeat.
FT DOMAIN 1 40 CISTERNAL SIDE (POTENTIAL).
FT TRANSMEM 41 61 POTENTIAL.
FT DOMAIN 62 1229 PORE SIDE (POTENTIAL).
FT DOMAIN 4 10 POLY-ALA.
FT DOMAIN 51 56 POLY-LYS.
FT DOMAIN 294 299 POLY-SER.
FT DOMAIN 441 444 POLY-PRO.
FT DOMAIN 499 502 POLY-SER.
FT DOMAIN 733 736 POLY-SER.
FT DOMAIN 819 826 POLY-SER.
FT DOMAIN 869 875 POLY-THR.
FT DOMAIN 1061 1067 POLY-SER.
FT CONFLICT 297 297 K -> E (IN REF. 2).
```

```

FT CONFLICT 336 336 S -> N (IN REF. 2)
FT CONFLICT 379 379 T -> A (IN REF. 2)
FT CONFLICT 451 451 K -> R (IN REF. 2)
FT CONFLICT 545 545 T -> I (IN REF. 2)
FT CONFLICT 559 559 T -> P (IN REF. 2)
FT CONFLICT 702 702 S -> P (IN REF. 2)
FT CONFLICT 881 881 P -> H (IN REF. 2)
FT CONFLICT 965 965 T -> A (IN REF. 2)
FT CONFLICT 982 984 PAT -> A (IN REF. 2)
FT CONFLICT 991 996 ASTIKI -> PSMIKV (IN REF. 2)
FT CONFLICT 1000 1000 H -> Y (IN REF. 2)
FT CONFLICT 1006 1008 OPT -> HPI (IN REF. 2)
FT CONFLICT 1044 1044 MISSING (IN REF. 2)
FT CONFLICT 1102 1102 A -> T (IN REF. 2)
FT CONFLICT 1124 1124 T -> A (IN REF. 2)
FT CONFLICT 1146 1146 S -> G (IN REF. 1)
FT CONFLICT 1165 1165 L -> Q (IN REF. 1)
FT CONFLICT 1195 1195 A -> G (IN REF. 1)
FT CONFLICT 1199 1229 SAAPSFICAGSKTPGARQLQARRQHTKK -> NTFAHQ
FT CONFLICT 1202 1202 QEHSRKGPNLXKRLPAVRAQGPFRGQASSFFTRKE
FT CONFLICT 1229 1202 P -> L (IN REF. 3)
SQ SEQUENCE 1229 AA; 125087 MW; 515655D1285898B CRC64;

Query Match 39.4%; Score 50; DB 1; Length 1229;
Best Local Similarity 57.9%; Pred. No. 1.4e+02;
Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

QY 1 PSTRPAGST--APPAHGV 17
DB 153 PPARPAPRSTPPSPPTHRV 171

RESULT 26
YL14_HUMAN
ID YL14_HUMAN STANDARD; PRT; 259 AA.
AC P40222; Q8N2Y3;
DT 01-FEB-1995 (Rel. 31, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein initially thought to be identical with
DE interleukin-14 (IL-14) (High molecular weight B-cell growth factor)
DE (HMW-BCGF).
DE IL14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93317675; PubMed=8327514;
RA Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA Tamayo A., Claypool K., McCourt D., Srikiatchatoorn A., Ford R.J.;
RT "Identification of a cDNA for a human high-molecular-weight B-cell
RT growth factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6330-6334(1993).
RN [2]
RP ERRATUM.
RX MEDLINE=96353961; PubMed=8755619;
RA Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA Tamayo A., Claypool K., McCourt D., Srikiatchatoorn A., Ford R.J.;
RL Proc. Natl. Acad. Sci. U.S.A. 93:8154-8154(1996).
RN [3]
RP SEQUENCE OF 130-259 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,

```

```

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnersch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- CAUTION: The clone described in Ref.1 was initially thought to
CC code for a high molecular weight interleukin (IL-14). The
CC translation of this cDNA does not predict the open reading frame
CC for the 60 kDa protein described in Ref.1. A reading frame on the
CC plus strand predicts a 7.7 kDa protein. The longest open reading
CC frame (shown here) is on the opposite strand and predicts a 36.4
CC kDa protein. The relationship of this sequence to IL-14, if any,
CC is uncertain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L15344; -; NOT ANNOTATED_CDS.
CC EMBL; BC029686; AAH29686.1; -.
CC PIR; A48203; A48203.
CC Genew; HGNC:5976; IL14.
CC MIM; 147684; -.
CC GO; GO:0005576; C:extracellular; NAS.
CC GO; GO:0030372; F:high molecular weight B-cell growth factor . . . ; NAS.
CC GO; GO:0008283; P:cell proliferation; NAS.
CC KW Hypothetical protein.
SQ SEQUENCE 259 AA; 30107 MW; 8AB8D1AA6891357C CRC64;

Query Match 39.0%; Score 49.5; DB 1; Length 259;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 11; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 PDTRP-APGSTAPPAGHVTAP 21
DB 214 PERPEGGAQAPSSPRVTEAP 235

RESULT 27
RTN1_HUMAN
ID RTN1_HUMAN STANDARD; PRT; 776 AA.
AC Q16799; Q16800; Q16801;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Reticulon 1 (Neuroendocrine-specific protein).
GN RTN1 OR NSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS RTN1-A; RTN1-B AND RTN1-C).
RC TISSUE=Lung carcinoma;
RX MEDLINE=93293865; PubMed=7685762;
RA Roebroek A.J.M., Van de Velde H.J.K., Van Bokhoven A., Broers J.L.V.,
RA Ramaekers F.C.S., Van de Ven W.J.M.;
RT "Cloning and expression of alternative transcripts of a novel
RT neuroendocrine-specific gene and identification of its 135-kDa
RT translational product.";
RL J. Biol. Chem. 268:13439-13447(1993).

```

RN [2]  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=96429995; PubMed=8833145;  
RA Roebroek A.J.M., Ayoubi T.A.Y., Van de Velde H.J.K.,  
RA Schoenmakers E.F.P.M., Pauli I.G.L., Van de Ven W.J.M.;  
RT "Genomic organization of the human NSP gene, prototype of a novel gene  
family encoding reticulons.";  
RL Genomics 32:191-199(1996).  
RN [3]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=9828245; PubMed=9560466;  
RA Hens J., Nuydens R., Geerts H., Senden N.H., Van de Ven W.J.M.,  
RA Roebroek A.J., van de Velde H.J.K., Ramaekers P.C., Broers J.L.;  
RT "Neuronal differentiation is accompanied by NSP-C expression.";  
RL Cell Tissue Res. 292:229-237(1998).  
CC -!- FUNCTION: May be involved in neuroendocrine secretion or in  
CC membrane trafficking in neuroendocrine cells.  
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=RTN1-A; Synonyms=NSP-A;  
CC IsoId=Q16799-1; Sequence=Displayed;  
CC Name=RTN1-B; Synonyms=NSP-B;  
CC IsoId=Q16799-2; Sequence=VSP\_005644;  
CC Name=RTN1-C; Synonyms=NSP-C;  
CC IsoId=Q16799-3; Sequence=VSP\_005645, VSP\_005646;  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN NEURAL AND NEUROENDOCRINE TISSUES  
CC AND CELL CULTURES DERIVED THEREFROM. EXPRESSION OF ISOFORM RTN1-C  
CC IS STRONGLY CORRELATED WITH NEURONAL DIFFERENTIATION.  
CC -!- PTM: Isoforms RTN1-A and RTN1-B are phosphorylated.  
CC -!- SIMILARITY: Contains 1 reticulon domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; L10333; AAA59950.1; -  
CC EMBL; L10334; AAA59951.1; -  
CC EMBL; L10335; AAA59952.1; -  
CC PIR; A46583; A46583.  
CC PIR; I60904; I60904.  
CC Gerew; HGNC:10467; RTN1.  
CC MIM; 60865; -  
CC GO; GO:0030176; C:integral to endoplasmic reticulum membrane; TAS.  
CC GO; GO:0004871; F:signal transducer activity; NAS.  
CC GO; GO:0030182; P:neuron differentiation; TAS.  
CC GO; GO:0007165; P:signal transduction; NAS.  
CC InterPro; IPR003388; Reticulon.  
CC Pfam; PF02453; Reticulon; 1.  
CC PROSITE; PS50845; RETICULON; 1.  
CC Endoplasmic reticulum; Alternative splicing; Transmembrane;  
KW Phosphorylation.  
FT TRANSMEM 603 623 POTENTIAL.  
FT TRANSMEM 726 746 POTENTIAL.  
FT DOMAIN 589 776 RETICULON.  
FT DOMAIN 609 612 POLY-LEU.  
FT VARSPPLIC 1 420 Missing (in isoform RTN1-B).  
FT /FTId=VSP\_005644.  
FT VARSPPLIC 1 568 Missing (in isoform RTN1-C).  
FT /FTId=VSP\_005645.  
FT VARSPPLIC 569 588 GEGPLGPGAPPPLFLINKQK -> MQATADSTRMDCVWSNW  
FT /FTId=VSP\_005646.  
FT KSK (in isoform RTN1-C).  
SQ SEQUENCE 776 AA; 83617 MW; CA5B6232353096FE CRC64;  
  
Query Match 39.0%; Score 49.5; DB 1; Length 776;  
Best Local Similarity 43.5%; Pred. No. 1e+02;  
Matches 10; Conservative 1; Mismatches 11; Indels 1; Gaps 1;

QY 1 PDTRPAPGSTAPPAGHGVTSAPDT 23  
Db 526 PSTEPQPGPELPPGDALE-PET 547  
  
RESULT 28  
MOBC\_THIFE STANDARD; PRT; 118 AA.  
AC P22899;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE MOBC protein.  
GN MOBC.  
OS Thiobacillus ferrooxidans.  
OC Plasmid pTF-FC2.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;  
OC Acidithiobacillaceae; Acidithiobacillus.  
OX NCBI\_TaxID=923;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93015664; PubMed=1400173;  
RA Rohrer J., Rawlings D.E.;  
RT "Sequence analysis and characterization of the mobilization region of  
RT a broad-host-range plasmid, pTF-FC2, isolated from Thiobacillus  
RT ferrooxidans.";  
RL J. Bacteriol. 174:6230-6237(1992).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; M57717; AAA27390.1; -  
CC PIR; C43256; C43256.  
CC Plasmid.  
SQ SEQUENCE 118 AA; 12956 MW; 13EAA3B1F1E707DE CRC64;  
  
Query Match 38.6%; Score 49; DB 1; Length 118;  
Best Local Similarity 55.6%; Pred. No. 18;  
Matches 10; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
  
QY 6 APGSTAPPAGHGVTSAPDT 23  
Db 87 APGDTPPPPAPAVKPADT 104  
  
RESULT 29  
HYBP\_AZOVI STANDARD; PRT; 303 AA.  
ID HYBP\_AZOVI  
AC P31880;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hydrogenase nickel incorporation protein hypB.  
GN HYBP.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=UM;  
RX MEDLINE=92305060; PubMed=1610901;  
RA Chen J.C., Mortenson L.E.;  
RT "Identification of six open reading frames from a region of the  
RT Azotobacter vinelandii genome likely involved in dihydrogen  
RT metabolism.";  
RL Biochim. Biophys. Acta 1131:199-202(1992).  
CC -!- FUNCTION: Could be involved in nickel binding and accumulation.



```

CC CC      -!- SIMILARITY: Belongs to the hypB/hupM family.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC DR      EMBL; X63650; CAA45184.1; -
CC DR      EMBL; L23970; AAI19509.1; -
CC DR      PIR; S23440; S23440.
CC DR      InterPro; IPR004392; HypB.
CC DR      InterPro; IPR002894; HypB_Ureg.
CC DR      Pfam; PF01495; HypB_Ureg; 1.
CC DR      TIGRFAMs; TIGR00073; hypB; 1.
CC KW      Metal-binding; Nickel.
CC FT      DOMAIN 18 30 HIS-RICH.
CC SQ      SEQUENCE 303 AA; 33185 MW; 743E6773527D1922 CRC64;

Query Match      38.6%; Score 49; DB 1; Length 303;
Best Local Similarity 69.2%; Pred. No. 46;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      4 RPAPGSTAPPAGH 16
DB      35 RPAPAEAAPPAAG 47

RESULT 30
HXB2_HUMAN
ID      HXB2_HUMAN STANDARD; PRT; 356 AA.
AC      P14652; P10913; P17485;
DT      01-APR-1990 (Rel. 14, Created)
DT      01-APR-1993 (Rel. 14, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Homeobox protein Hox-B2 (Hox-2H) (Hox-2.8) (K8).
GN      HOXB2 OR HOX2H.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90098876; PubMed=2574852;
RA      Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,
RA      Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
RT      "The human HOX gene family.";
RL      Nucleic Acids Res. 17:10385-10402(1989).
RN      [2]
RP      SEQUENCE OF 132-208 FROM N.A.
RC      TISSUE=Placenta;
RX      MEDLINE=89378558; PubMed=2570724;
RA      Giampaolo A., Acampora D., Zappavigna V., Pannese M.,
RA      D'Esposito M., Care A., Faiella A., Stornaiuolo A., Russo G.,
RA      Simeone A., Boncinelli E., Peschle C.;
RT      "Differential expression of human HOX-2 genes along the anterior-
RT      posterior axis in embryonic central nervous system.";
RL      Differentiation 40:191-197(1989).
RN      [3]
RP      SEQUENCE OF 143-208 FROM N.A.
RX      MEDLINE=90215256; PubMed=2576652;
RA      Boncinelli E., Acampora D., Pannese M., D'Esposito M., Somma R.,
RA      Gaudino G., Stornaiuolo A., Cafiero M., Faiella A., Simeone A.;
RT      "Organization of human class I homeobox genes.";
RL      Genome 31:745-756(1989).
RN      [4]
RP      SEQUENCE OF 143-202 FROM N.A.
RX      MEDLINE=88329001; PubMed=2901346;
RA      Kongsuwan K., Webb E., Housiaux F., Adams J.M.;
RT      "Expression of multiple homeobox genes within diverse mammalian
RT      haemopoietic lineages.";

```

```

RL      EMBO J. 7:2131-2138(1988).
RN      [5]
RX      SEQUENCE OF 1-42 FROM N.A.
RX      MEDLINE=95181447; PubMed=7876223;
RA      Vieille-Grosjean I., Huber P.;
RT      "Transcription factor GATA-1 regulates human HOXB2 gene expression in
RT      erythroid cells.";
RL      J. Biol. Chem. 270:4544-4550(1995).
CC      -!- FUNCTION: Sequence-specific transcription factor which is part of
CC      a developmental regulatory system that provides cells with
CC      specific positional identities on the anterior-posterior axis.
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- DEVELOPMENTAL STAGE: Expressed in whole embryos and fetuses at 5-9
CC      weeks from conception.
CC      -!- SIMILARITY: BELONGS TO THE ANTP HOMEODOMAIN FAMILY.
CC      PROBOSCIPEDIA SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC DR      EMBL; X16665; CAA34655.1; -
CC DR      EMBL; X16176; CAA34298.1; -
CC DR      EMBL; X14571; CAA32709.1; -
CC DR      EMBL; X78978; CAA55581.1; -
CC DR      PIR; S07542; WJH02H.
CC DR      HSSP; P14653; 1B72.
CC DR      TRANSFAC; T03323; -
CC DR      Genew; HGNC:5113; HOXB2.
CC DR      MIM; 142967; -
CC DR      GO; GO:0008015; P:circulation; TAS.
CC DR      GO; GO:0007275; P:development; TAS.
CC DR      InterPro; IPR001827; Antennapedia.
CC DR      InterPro; IPR001356; Homeobox.
CC DR      InterPro; IPR000047; HTH_lambdarepress.
CC DR      Pfam; PF00046; homeobox; 1.
CC DR      PRINTS; PR00025; ANTENNAPEDIA.
CC DR      PRINTS; PR00024; HOMEODOMAIN.
CC DR      PRINTS; PR00031; HTHREPRESS.
CC DR      ProDom; PD000010; Homeobox; 1.
CC DR      SMART; SM00389; HOX; 1.
CC DR      PROSITE; PS00027; HOMEODOMAIN; 1.
CC DR      PROSITE; PS00032; ANTENNAPEDIA; 1.
CC DR      PROSITE; PS00071; HOMEODOMAIN; 2; 1.
CC KW      Homeobox; DNA-binding; Developmental protein; Nuclear protein;
CC KW      Transcription regulation.
FT      SITE 94 99 ANTP-TYPE HEXAPEPTIDE.
FT      DNA_BIND 143 202 HOMEODOMAIN.
FT      CONFLICT 136 138 PEA -> RRL (IN REF. 2).
SQ      SEQUENCE 356 AA; 37913 MW; D7ACA0E3D5BEBFB9 CRC64;

Query Match      38.6%; Score 49; DB 1; Length 356;
Best Local Similarity 50.0%; Pred. No. 54;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY      3 TRPAPGSTAPPAGHGVTSAPD 22
DB      112 TSPSPAASAVPASGVGSPAD 131

RESULT 31
APA_MYCAV
ID      APA_MYCAV STANDARD; PRT; 381 AA.
AC      Q48919;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Alanine and proline-rich secreted protein apa precursor (45/47 kDa
DE      antigen) (Fibronectin attachment protein) (FAP-A).

```



```
GN APA.
OS Mycobacterium avium.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=101;
RX MEDLINE=97011577; PubMed=8858587;
RA Schorey J.S., Holsti M.A., Ratliff T.L., Allen P.M., Brown E.J.;
RT *Characterization of the fibronectin-attachment protein of
RT Mycobacterium avium reveals a fibronectin-binding motif conserved
RT among mycobacteria."
RL Mol. Microbiol. 21:321-329(1996).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U53585; AAB50543.1; -.
DR Antigen; Signal.
KW SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 381 ALANINE AND PROLINE-RICH SECRETED
FT PROTEIN APA.
FT SEQUENCE 381 AA; 38210 MW; 0DACB9A22AA1D2B CRC64;
Query Match 38.6%; Score 49; DB 1; Length 381;
Best Local Similarity 52.3%; Pred. No. 58;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PAPGSTAPPAGHGVTSAP 21
Db 317 PAPGAPGAPAGAPAPAP 333
||||: ||| :||

RESULT 32
VGLZ HSVK STANDARD; PRT; 383 AA.
AC P32515;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glycoprotein precursor.
GN US4.
OS Equine herpesvirus type 1 (strain Kentucky A) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92263758; PubMed=1316673;
RA Colle C.F. III, Flowers C.C., O'Callaghan D.J.;
RT "Open reading frames encoding a protein kinase, homolog of
RT glycoprotein gX of pseudorabies virus, and a novel glycoprotein map
RT within the unique short segment of equine herpesvirus type 1."
RL Virology 188:545-557(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M87497; AAA46072.1; -.
DR PIR; C42538; VGBEKG.
DR InterPro; IPR007110; Ig-like.
```

```
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 383 GLYCOPROTEIN.
FT DOMAIN 23 75 SER/THR-RICH.
FT TRANSMEM 354 371 POTENTIAL.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 383 AA; 41027 MW; B390611414131C2B CRC64;

Query Match 38.6%; Score 49; DB 1; Length 383;
Best Local Similarity 43.5%; Pred. No. 59;
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 1 PDTEPAPGSTAPPAGHGVTSAPDT 23
Db 53 PTTSPTTSSSPTSTHTSSPST 75
||||: ||| :|||

RESULT 33
HMES DROME
ID HMES DROME STANDARD; PRT; 497 AA.
AC P18488; Q9VFC1;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Empty spiracles homeotic protein.
DE EMS OR E4 OR CG2988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=92289690; PubMed=1376248;
RA Walldorf U., Gehring W.J.;
RT "Empty spiracles, a gap gene containing a homeobox involved in
RT Drosophila head development."
RL EMBO J. 11:2247-2259(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=90152339; PubMed=2576012;
RA Dalton D., Chadwick R., McGinnis W.;
RT "Expression and embryonic function of empty spiracles: a Drosophila
RT homeo box gene with two patterning functions on the
RT anterior-posterior axis of the embryo."
RL Genes Dev. 3:1940-1956(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Evans R., Fouts D.E., Garg N.S., Gilbert W.M., Glasser K.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kaush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
```





FT TURN 173 174  
 FT STRAND 183 191  
 FT TURN 192 193  
 FT STRAND 194 201  
 FT HELIX 211 213  
 FT HELIX 214 221  
 FT STRAND 224 224  
 FT TURN 225 228  
 FT STRAND 229 229  
 FT TURN 231 232  
 FT HELIX 237 245  
 FT HELIX 247 249  
 FT HELIX 250 257  
 FT HELIX 259 261  
 FT STRAND 263 263  
 FT HELIX 266 268  
 FT STRAND 271 272  
 FT STRAND 277 280  
 FT TURN 282 283  
 FT STRAND 286 295  
 FT TURN 296 297  
 FT STRAND 301 301  
 FT HELIX 304 313  
 FT TURN 315 316  
 FT HELIX 317 327  
 FT TURN 328 329  
 FT STRAND 333 342  
 FT HELIX 343 345  
 FT TURN 346 347  
 FT TURN 353 354  
 FT TURN 356 357  
 FT TURN 362 364  
 FT STRAND 368 377  
 FT HELIX 382 385  
 FT TURN 386 388  
 FT TURN 393 394  
 FT TURN 398 399  
 FT STRAND 400 401  
 FT HELIX 406 423  
 FT TURN 424 424  
 FT TURN 426 427  
 FT HELIX 428 430  
 FT HELIX 432 434  
 FT TURN 448 449  
 FT STRAND 453 454  
 FT TURN 465 468  
 FT STRAND 471 471  
 FT STRAND 482 482  
 FT STRAND 488 495  
 FT HELIX 499 501  
 FT HELIX 506 513  
 FT TURN 514 514  
 FT HELIX 517 532  
 FT TURN 533 533  
 FT HELIX 537 548  
 FT TURN 549 550

Query Match 38.6%; Score 49; DB 1; Length 753;  
 Best Local Similarity 43.5%; Pred. No. 1.2e+02;  
 Matches 10; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHVTSPDPT 23  
 Db 38 PAABPTPGAGTAPGSLXAPDT 60

RESULT 36  
 CG2F MOUSE STANDARD; PRT; 776 AA.  
 ID P51944; Q60797; Q60799;  
 AC 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE G2/mitotic-specific cyclin F.  
 GN CCNF.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RX MEDLINE=95284479; PubMed=7767003;  
 RA Obermayr F.O., Sutherland H.F., Kraus B., Frischauf A.-M.;  
 RT "Mouse cyclin F maps to a conserved linkage group on mouse chromosome  
 17.";  
 RL Mamm. Genome 6:149-150(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
 RA Croci L., Stayton C.L., Bossolasco M., Bianchi E., Corradi A.M.,  
 RA Pardi R., Consalez G.G.;  
 RT "Expression of cyclin F at early stages of mouse embryonic brain  
 development.";  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Likely to be involved in the control of the cell cycle  
 CC during S phase and G2 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Short;  
 CC IsoId=P51944-1; Sequence=Displayed;  
 CC Name=Long;  
 CC IsoId=P51944-2; Sequence=VSP\_001257;  
 CC -!- SIMILARITY: Belongs to the cyclin family. Cyclin AB subfamily.  
 CC -!- SIMILARITY: Contains 1 F-box domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z47766; CAA87695.1; -.  
 DR EMBL; U20612; AAA62317.1; -.  
 DR EMBL; U20636; AAA63152.1; -.  
 DR PIR; I48317; I48317.  
 DR HSSP; P30274; IVIN.  
 DR MGD; MGI:102551; Ccnf.  
 DR InterPro; IPR006670; Cyclin.  
 DR InterPro; IPR004367; Cyclin Cterm.  
 DR InterPro; IPR006671; Cyclin\_N.  
 DR InterPro; IPR001810; F-box.  
 DR Pfam; PF00134; cyclin; 1.  
 DR Pfam; PF02984; cyclin C; 1.  
 DR Pfam; PF00646; F-box; 1.  
 DR SMART; SM00385; CYCLIN; 2.  
 DR SMART; SM00256; FBOX; 1.  
 DR PROSITE; PS00252; CYCLINS; 1.  
 DR PROSITE; PS00181; FBOX; 1.  
 KW Cyclin; Cell cycle; Cell division; Mitosis; Nuclear protein;  
 KW Alternative splicing.  
 FT DOMAIN 29 76  
 FT VARSPLIC 5 5  
 FT F-BOX.  
 FT G -> GGECWRDGVGDRSYGTEAGGAWTRAPVTRCASHPP  
 FT KHGV (in isoform Long).  
 FT /FTID=VSP\_001257.  
 FT L -> P (IN REF. 2).  
 FT Q -> E (IN REF. 2).  
 FT T -> S (IN REF. 2).  
 FT A -> T (IN REF. 2).  
 FT S -> T (IN REF. 2).  
 FT G -> V (IN REF. 2).  
 FT T -> S (IN REF. 2).  
 FT A -> S (IN REF. 2).  
 FT H -> R (IN REF. 2).  
 FT Q -> R (IN REF. 2).

```
FT CONFLICT 567 568 GE -> ERR (IN REF. 2).
FT CONFLICT 602 602 S -> G (IN REF. 2).
FT CONFLICT 623 623 E -> R (IN REF. 2).
SQ SEQUENCE 776 AA; 86320 MW; 786399381180B993 CRC64;

Query Match 38.6%; Score 49; DB 1; Length 776;
Best Local Similarity 64.7%; Pred. No. 1.2e+02;
Matches 11; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

QY 5 PAPGSTAPPAGHVTISAP 21
Db 665 PAPGTQAPPA----SAP 677

RESULT 37
DIAL_MOUSE STANDARD; PRT; 1255 AA.
AC O08808;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Diaphanous protein homolog 1 (Diaphanous-related formin 1) (DRF1)
DE (mDial) (p140mDia).
GN DIAPH1 OR DIAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97357293; PubMed=9214622;
RA Watanabe N., Madaule P., Reid T., Ishizaki T., Watanabe G.,
RA Kakizuka A., Saito Y., Nakao K., Jockusch B.M., Narumiya S.;
RT "p140mDia, a mammalian homolog of Drosophila diaphanous, is a target
RT protein for Rho small GTPase and is a ligand for profilin.";
RL EMBO J. 16:3044-3056(1997).
RN [2]
RP FUNCTION.
RX MEDLINE=20142655; PubMed=10678165;
RA Tomimaga T., Sahai E., Chardin P., McCormick F., Courtneidge S.A.,
RA Alberts A.S.;
RT "Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase
RT signaling.";
RL Mol. Cell 5:13-25(2000).
CC -!- FUNCTION: Binds to GTP-bound form of Rho and to profilin. Acts in
CC a Rho-dependent manner to recruit profilin to the membrane, where
CC it promotes actin polymerization. It is required for cytokinesis,
CC stress fiber formation, and transcriptional activation of the
CC serum response factor. DFR proteins couple Rho and Src tyrosine
CC kinase during signaling and the regulation of actin dynamics.
CC -!- SUBCELLULAR LOCATION: MEMBRANE RUFFLES, ESPECIALLY AT THE TIP OF
CC RUFFLES, OF MOTILE CELLS.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- DOMAIN: DRFs are regulated by intramolecular GBD-DAD binding where
CC Rho-GTP activates the DRFs by disrupting the GBD-DAD interaction.
CC -!- SIMILARITY: Contains 1 GTPase-binding (GBD) domain.
CC -!- SIMILARITY: Contains 1 Formin homology 1 (FH1) domain.
CC -!- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.
CC -!- SIMILARITY: Contains 1 Formin homology 3 (FH3) domain.
CC -!- SIMILARITY: Contains 1 DRF autoregulatory (DAD) domain.
CC -!- SIMILARITY: Belongs to the formin homology family. Diaphanous
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U96963; AAC53280.1; -.
CC PIR; T31065; T31065.
```

```
DR MGD; MGI:1194490; Diapl.
DR GO; GO:0005515; F:protein binding; IPI.
DR InterPro; IPR003104; FH2.
DR Pfam; PF02181; FH2; 1.
DR SMART; SM00498; FH2; 1.
KW Coiled coil; Repeat.
FT DOMAIN 460 562 COILED COIL (POTENTIAL).
FT DOMAIN 63 260 GBD.
FT DOMAIN 157 457 FH3.
FT DOMAIN 586 747 FH1 (PRO-RICH).
FT DOMAIN 752 1197 FH2.
FT DOMAIN 1027 1179 COILED COIL (POTENTIAL).
FT DOMAIN 1180 1194 DAD.
FT DOMAIN 1196 1199 ARG/LYS-RICH (BASIC).
SQ SEQUENCE 1255 AA; 139343 MW; 09404164873CA7C1 CRC64;

Query Match 38.6%; Score 49; DB 1; Length 1255;
Best Local Similarity 44.0%; Pred. No. 1.9e+02;
Matches 11; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 1 PDTRPAPGSTA---PPAHGVTISAP 21
Db 648 PPPPLPGATAIPPPPLPGATAIP 672

RESULT 38
NK32_HUMAN STANDARD; PRT; 333 AA.
AC P78367;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein Nkx-3.2 (Bagpipe homeobox protein homolog 1).
GN BAPX1 OR NKX3H.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98086223; PubMed=9426254;
RA Tribioli C., Lufkin T.;
RT "Molecular cloning, chromosomal mapping and developmental expression
RT of BAPX1, a novel human homeobox-containing gene homologous to
RT Drosophila bagpipe.";
RL Gene 203:225-233(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98089336; PubMed=9344671;
RA Yoshiura K.I., Murray J.C.;
RT "Sequence and chromosomal assignment of human BAPX1, a bagpipe-related
RT gene, to 4p16.1: a candidate gene for skeletal dysplasia.";
RL Genomics 45:425-428(1997).
RN [3]
RP SEQUENCE OF 206-265 FROM N.A.
RX MEDLINE=97398454; PubMed=9256352;
RA Tribioli C., Frasch M., Lufkin T.;
RT "Bapx1: an evolutionary conserved homologue of the Drosophila bagpipe
RT homeobox gene is expressed in splanchnic mesoderm and the embryonic
RT skeleton.";
RL Mech. Dev. 65:145-162(1997).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Expressed in visceral mesoderm and embryonic
CC skeleton.
CC -!- SIMILARITY: Belongs to the NK-3 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```



```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF005260; AAC39536.1; -.
DR EMBL; AF003801; AAB82783.1; -.
DR EMBL; AF003802; AAB82784.1; -.
DR EMBL; U89845; AAB49696.1; -.
DR HSSP; P22808; 1VND.
DR TRANSFAC; T02668; -.
DR Genew; HGNC:951; BAPX1.
DR MIM; 602183; -.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR GO; GO:0001501; P:skeletal development; TAS.
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000047; HTH lambrapressr.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW Homeobox; DNA-binding; Nuclear protein.
FT DOMAIN 180 188 POLY-GLY.
FT DOMAIN 194 198 POLY-GLU.
FT DNA_BIND 206 265 HOMEBOX.
FT DOMAIN 327 330 POLY-ALA.
SQ SEQUENCE 333 AA; 34813 MW; 8C406E188D27780B CRC64;

Query Match 38.2%; Score 48.5; DB 1; Length 333;
Best Local Similarity 52.4%; Pred. No. 58;
Matches 11; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

QY 1 PDTRPAGSTAPAHGVTSAP 21
Db 30 PEGRPAGGTAA---SVAAAP 47

RESULT 39
DAG1_MOUSE
ID DAG1_MOUSE STANDARD; PRT; 893 AA.
AC Q62165; Q61094; Q61141; Q61497;
DT 01-NOV-1997 (Rel. 35, Created);
DT 16-OCT-2001 (Rel. 40, Last sequence update);
DT 15-MAR-2004 (Rel. 43, Last annotation update);
DE Dystroglycan precursor (Dystrophin-associated glycoprotein 1)
DE [Contains: Alpha-dystroglycan (Alpha-DG); Beta-dystroglycan (Beta-
DE DG)].
GN DAG1 OR DAG-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1] _TaxID=10090;
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=129/SvJ;
RX MEDLINE=97318791; PubMed=9175728;
RA Williamson R.A., Henry M.D., Daniels K.J., Hrstka R.F., Lee J.C.,
RA Sunada Y., Ibraghimov-Beskrovnaya O., Campbell K.P.;
RT "Dystroglycan is essential for early embryonic development:
RT disruption of Reichert's membrane in Dag1-null mice.";
RL Hum. Mol. Genet. 6:831-841(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
```

```
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16893-16903(2002).
RN [3]
RP SEQUENCE OF 1-650 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=97210754; PubMed=9057818;
RA Brancaccio A., Ruegg M.A., Engel J.;
RT "Cloning and sequencing of mouse skeletal muscle alpha-dystroglycan.";
RL Matrix Biol. 14:681-685(1995).
RN [4]
RP REVISIONS TO 142-143.
RA Brancaccio A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 352-650 FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/10; TISSUE=Skeletal muscle;
RX MEDLINE=95135414; PubMed=7833916;
RA Gorecki D.C., Derry J.M.J., Barnard E.A.;
RT "Dystroglycan: brain localisation and chromosome mapping in the
RT mouse.";
RL Hum. Mol. Genet. 3:1589-1597(1994).
RN [6]
RP SEQUENCE OF 620-893 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Decidua;
RX MEDLINE=97026286; PubMed=8872465;
RA Yotsumoto S., Fujiwara H., Horton J.H., Mosby T.A., Wang X.,
RA Cui Y., Ko M.S.H.;
RT "Cloning and expression analyses of mouse dystroglycan gene: specific
RT expression in maternal decidua at the peri-implantation stage.";
RL Hum. Mol. Genet. 5:1259-1267(1996).
RN [7]
RP DISULFIDE BOND.
RX MEDLINE=99116369; PubMed=9917844;
RA Brancaccio A., Jeno P., Engel J.;
RT "A single disulfide bridge (Cys182-Cys264) is crucial for alpha-
RT dystroglycan N-terminal domain stability.";
RL Ann. N.Y. Acad. Sci. 857:228-231(1998).
CC -!- FUNCTION: Forms part of the dystrophin-associated protein complex
CC (DAPC) which may link the cytoskeleton to the extracellular
CC matrix. Alpha-dystroglycan functions as a laminin receptor. Binds
CC to several types of arenaviruses. Is a target for the entry of
CC Mycobacterium leprae into peripheral nerve Schwann cells.
CC -!- SUBCELLULAR LOCATION: Alpha-dystroglycan is an extracellular
CC protein, while beta-dystroglycan is a type-I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in a variety of tissues. In brain,
CC expressed in the hippocampal formation, the olfactory bulb, the
CC cerebellum and the thalamus.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U48854; AAA99779.2; -.
CC EMBL; BC007150; AAH07150.1; -.
CC EMBL; X86073; CAA60031.1; -.
CC EMBL; Z34532; CAA84293.1; -.
CC EMBL; U43512; AAC52853.1; -.
CC PIR; S59630; S59630.
```

```
DR MGD; MGI:101864; Dag1.
DR GO; GO:0016011; C:dystroglycan complex; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
DR InterPro; IPR006644; Cadg.
DR InterPro; IPR008465; DAG1.
DR Pfam; PF05454; DAG1; 1.
DR SMART; SM00736; CADG; 2.
KW Signal; Glycoprotein; Transmembrane; Cytoskeleton.
FT SIGNAL 1 27
FT CHAIN 28 651
FT CHAIN 652 893
FT DOMAIN 652 751
FT TRANSMEM 752 772
FT DOMAIN 773 893
FT DOMAIN 316 475
FT DOMAIN 807 893
FT DISULFID 180 262
FT DISULFID 667 711
FT CARBOHYD 139 139
FT CARBOHYD 483 483
FT CARBOHYD 639 639
FT CARBOHYD 647 647
FT CARBOHYD 659 659
FT CONFLICT 448 450
FT CONFLICT 599 600
FT CONFLICT 643 643
FT SEQUENCE 893 AA; 96904 MW; 59C081EA86AB0AC1 CRC64;

Query Match 38.2%; Score 48.5; DB 1; Length 893;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 11; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22
Db 718 EVAPPSPGSSAAPA---TEVPD 736

RESULT 40
IF2_STRCO
ID IF2_STRCO STANDARD; PRT; 1033 AA.
AC Q8CJQ8;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Translation initiation factor IF-2.
GN INF8 OR SC05706 OR SC9F2.10C OR SC5H4.30.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and protects its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
```

```
CC -!- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL939124; CAD55362.1; -
CC PIR; T35989; T35989.
CC HAMAP; MF 00100; -; 1.
CC InterPro; IPR000637; AT_hook.
CC InterPro; IPR001851; Bac_inmem_transp.
CC InterPro; IPR004161; EFTU_D2.
CC InterPro; IPR000795; EF_GTPbind.
CC InterPro; IPR000178; IF2.
CC InterPro; IPR006847; IF2_N.
CC InterPro; IPR001806; Ras_tmstrmng.
CC InterPro; IPR005225; Small_GTP.
CC InterPro; IPR009000; Translat_factor.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC Pfam; PF04760; IF2_N; 2.
CC PRINTS; PR00929; ATHOOK.
CC PRINTS; PR00315; ELONGATNFCT.
CC PRINTS; PR00449; RASTRNSFRMNG.
CC ProDom; PD186100; IF2; 1.
CC TIGRFAMs; TIGR00487; IF-2; 1.
CC TIGRFAMs; TIGR00231; small_GTP; 1.
CC PROSITE; PS01176; IF2; FALSE_NEG.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 529 681 G-DOMAIN.
FT NP_BIND 535 542 GTP (BY SIMILARITY).
FT NP_BIND 585 589 GTP (BY SIMILARITY).
FT NP_BIND 639 642 GTP (BY SIMILARITY).
SQ SEQUENCE 1033 AA; 105657 MW; 47D581FB0072A045 CRC64;

Query Match 38.2%; Score 48.5; DB 1; Length 1033;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 13; Conservative 2; Mismatches 6; Indels 5; Gaps 2;

QY 1 PDTRPAPGS---TAPPAGHGVTSAPDT 23
Db 133 PAPRPAPAPAEFTAPPA--APAAPST 156
```

Search completed: May 6, 2004, 16:29:41  
Job time : 13.3846 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 16:25:44 ; Search time 43.0513 Seconds  
(without alignments)  
168.565 Million cell updates/sec

Title: US-10-070-566-7  
Perfect score: 127  
Sequence: 1 PDTRPAPGSTAPPAGHGVTSAPDT 23

Scoring table: 3LOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

1:	sp_archaea:*
2:	sp_bacteria:*
3:	sp_fungi:*
4:	sp_human:*
5:	sp_invertebrate:*
6:	sp_mammal:*
7:	sp_mhc:*
8:	sp_organelle:*
9:	sp_phage:*
10:	sp_plant:*
11:	sp_rodent:*
12:	sp_virus:*
13:	sp_vertebrate:*
14:	sp_unclassified:*
15:	sp_rvirus:*
16:	sp_bacteriap:*
17:	sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	109	85.8	20	4	Q9UMI8	Q9umi8 homo sapien
2	92	72.4	553	6	Q9MZL1	Q9mzl1 macaca mula
3	63	49.6	564	10	Q8LIP8	Q8lip8 oryza sativ
4	61	48.0	1039	5	Q9VJ66	Q9vj66 drosophila
5	61	48.0	1334	16	Q9RKR9	Q9rkr9 streptomyce
6	59	46.5	306	10	Q9LHY8	Q9lhy8 oryza sativ
7	59	46.5	433	4	Q9UHM8	Q9uhm8 homo sapien
8	59	46.5	481	4	Q9NZ07	Q9nzo7 homo sapien
9	59	46.5	484	4	Q9UMX1	Q9umx1 homo sapien
10	59	46.5	484	4	Q9UHK2	Q9uhk2 homo sapien
11	59	46.5	652	3	Q99109	Q99109 ustilago ma
12	58	45.7	704	6	O19130	O19130 canis famli
13	57.5	45.3	745	16	Q89X06	Q89x06 bradyrhizob
14	57	44.9	650	10	Q7XLQ4	Q7xlq4 oryza sativ
15	56.5	44.5	407	10	Q9LEY7	Q9ley7 arabidopsis
16	56.5	44.5	409	10	Q8L708	Q8l708 arabidopsis

17	56	44.1	411	16	Q9RD79	Q9rd79 streptomyce
18	56	44.1	508	5	Q9BLR4	Q9blr4 leishmania
19	56	44.1	544	10	Q7XZD2	Q7xzd2 raphanus sa
20	56	44.1	1895	5	Q9VWS5	Q9vws5 drosophila
21	55.5	43.7	569	16	Q9KGV9	Q9kgv9 listeria mo
22	55	43.3	130	9	Q9ZXJ0	Q9zxj0 bacterioph
23	55	43.3	160	6	Q8SQ36	Q8sq36 bos taurus
24	55	43.3	160	6	Q8SQ37	Q8sq37 bos taurus
25	55	43.3	160	6	Q8SQ38	Q8sq38 bos taurus
26	55	43.3	193	16	Q9F3H9	Q9f3h9 streptomyce
27	55	43.3	329	10	Q9SM15	Q9sm15 zea mays (m
28	55	43.3	554	6	Q9SL89	Q9sl89 bos taurus
29	55	43.3	580	6	Q8WML4	Q8wml4 bos taurus
30	55	43.3	602	6	O19115	O19115 oryctolagus
31	55	43.3	916	5	O45360	O45360 caenorhabdi
32	55	43.3	1339	4	Q9ULK3	Q9ulk3 homo sapien
33	54.5	42.9	509	5	Q8IQ93	Q8iq93 drosophila
34	54.5	42.9	656	5	Q9Y114	Q9y114 drosophila
35	54	42.5	85	4	Q8TE25	Q8te25 homo sapien
36	54	42.5	455	3	Q00881	Q00881 nectria hae
37	54	42.5	508	11	Q8CD55	Q8cd55 mus musculu
38	54	42.5	533	11	Q7TQE2	Q7tqe2 mus musculu
39	54	42.5	576	16	Q9XA04	Q9xa04 streptomyce
40	54	42.5	635	16	Q8NLV6	Q8nlv6 corynebacte
41	54	42.5	730	11	Q80XI8	Q80xi8 mus musculu
42	54	42.5	807	5	Q9VC14	Q9vc14 drosophila
43	54	42.5	889	16	Q9F2N5	Q9f2n5 streptomyce
44	54	42.5	905	4	Q7Z731	Q7z731 homo sapien
45	54	42.5	932	4	Q7Z2L2	Q7z2l2 homo sapien

ALIGNMENTS

RESULT 1

Q9UMI8  
ID Q9UMI8 PRELIMINARY; PRT; 20 AA.  
AC Q9UMI8;  
DT 01-MAY-2000 (TremBLrel. 13, Created)  
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)  
DE Mucin (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89235154; PubMed=2715633;  
RA King P.X., Tjandra J.J., Reynolds K., McLaughlin P.J., Purcell D.F.J.,  
RA McKenzie I.F.C.;  
RT "Reactivity of anti-human milk fat globule antibodies with synthetic  
RT peptides.";  
RL J. Immunol. 142:3503-3509(1989).  
DR EMBL; M26316; AAA36336.1; -.  
DR PIR; S10571; S10571.  
FT NON\_TER 1  
FT NON\_TER 20  
SQ SEQUENCE 20 AA; 1887 MW; 5B3473EAEBAFAD87 CRC64;

Query Match 85.8%; Score 109; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.6e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAGHGVTS A 20  
Db 1 PDTRPAPGSTAPPAGHGVTS A 20

RESULT 2  
Q9MZL1 PRELIMINARY; PRT; 553 AA.  
ID Q9MZL1  
AC Q9MZL1;

DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Mucin 1 (Fragment).  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20330533; PubMed=10869775;  
 RA Vaughan H.A., Ho D.W.M., Karanikas V., Sandrin M.S., McKenzie I.F.C.,  
 RA Pietersz G.A.;  
 RT "The Immune Response of Mice and Cynomolgus Monkeys to Macaque Mucin1-  
 RT Mannan.";  
 RL Vaccine 18:3297-3309(2000).  
 DR EMBL; AF176947; AAF82403.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR000082; SEA\_domain.  
 DR Pfam; PF01390; SEA; 1.  
 DR SMART; SM00200; SEA; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 DR PROSITE; PS50024; SEA; 1.  
 FT NON\_TER 1  
 FT NON\_TER 553  
 SQ SEQUENCE 553 AA; 55778 MW; 6D7E6DD2EE929318 CRC64;

Query Match 72.4%; Score 92; DB 6; Length 553;  
 Best Local Similarity 78.3%; Pred. No. 0.0011;  
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAPDT 23  
 ||| ||||| ||||| ||||| |||||  
 Db 100 PDTSAPGSGTPPARVTSAPDT 122

## RESULT 3

Q8LIP8  
 ID Q8LIP8 PRELIMINARY; PRT; 564 AA.  
 AC Q8LIP8;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Putative PRLI-interacting factor G.  
 GN OJ1370\_E02.13 OR OJ1354\_H07.23.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC  
 RT clone:OJ1370\_E02.13";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC  
 RT clone:OJ1354\_H07.23";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003756; BAC10056.1; -.  
 DR EMBL; AP003755; BAC21341.1; -.  
 DR Gramene; Q8LIP8; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR001005; Myb\_DNA\_binding.  
 DR PROSITE; PS00037; MYB 1; 1.  
 SQ SEQUENCE 564 AA; 61268 MW; F6D89CF602B5ADF0 CRC64;

Query Match 49.6%; Score 63; DB 10; Length 564;  
 Best Local Similarity 57.1%; Pred. No. 5.2;  
 Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 PDTRPAPGSTAPPAHGVTSAP 21  
 ||| ||||| ||||| ||||| |||||  
 Db 39 PDLRPPPPSLAPPATAASSP 59

## RESULT 4

Q9VJ66  
 ID Q9VJ66 PRELIMINARY; PRT; 1039 AA.  
 AC Q9VJ66;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE MSL-1 protein (Fragment).  
 GN MSL-1 OR CG10385.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman Y.-H.C., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Wan K.H., Doyle C., Rogers J., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Kvangalishia C.C., Perraz C., Ferreira S., Fleischmann W.,  
 RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., My M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003660; AAF53689.1; -.  
 DR FlyBase; FBgn0005617; msl-1.  
 DR GO; GO:0016456; C:dosage compensation complex (sensu Drosophila); NAS.  
 DR GO; GO:0003682; F:chromatin binding; IDA.  
 DR GO; GO:0009047; P:dosage compensation, by hyperactivation of . . .; NAS.  
 FT NON\_TER 1039  
 SQ SEQUENCE 1039 AA; 117503 MW; 644C21C0390A2750 CRC64;

Query Match 48.0%; Score 61; DB 5; Length 1039;  
 Best Local Similarity 59.1%; Pred. No. 17;  
 Matches 13; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

QY 1 PDTRPAGSTAPPAGHGVTSAPD 22  
 748 PSTTPTPGST--PQHAVTSSMD 767

Db

RESULT 5

Q9RKR9 PRELIMINARY; PRT; 1334 AA.

AC Q9RKR9

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative multi-domain regulatory protein.

GN SCO2259 OR SCC75A.05C.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Seeger K.J., Harris D.;

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mcl. Microbiol. 21:77-96(1996).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;

RX MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,

RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces

RT coelicolor A3(2)."

RL Nature 417:141-147(2002).

DR EMBL; AL39112; CAB61705.1; -.

DR PIR; T50568; T50568.

DR GO; GO:0003793; F:defense/immunity protein activity; IEA.

DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0000156; F:two-component response regulator activity; IEA.

DR GO; GO:0006952; P:defense response; IEA.

DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.

DR InterPro; IPR005158; BAD.

DR InterPro; IPR000767; Disease\_resist.

DR InterPro; IPR008941; TPR-like.

DR InterPro; IPR001867; Trans\_reg\_C.

DR Pfam; PF03704; BTAD; 1.

DR Pfam; PF00486; trans\_reg\_C; 1.

DR PRINTS; PR00364; DISEASERSIST.

DR ProDom; PD000329; Trans\_reg\_C; 1.

KW Complete proteome.

SQ SEQUENCE 1334 AA; 138787 MW; 78DC746883E8778C CRC64;

Query Match 48.0%; Score 61; DB 16; Length 1334;  
 Best Local Similarity 60.9%; Pred. No. 22;  
 Matches 14; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23  
 420 PGTAPAGST--PAGTVPAPGT 440

Db

RESULT 6

Q9LHY8 PRELIMINARY; PRT; 306 AA.

AC Q9LHY8

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ESTs D15336(C0474).

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI\_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC

RT clone:P0029D06.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP001552; BAA93036.1; -.

DR Gramene; Q9LHY8; -.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF00076; rrm; 1.

DR SMART; SM00360; RRM; 1.

DR PROSITE; PS0102; RRM; 1.

SQ SEQUENCE 306 AA; 32284 MW; F90911740846B540 CRC64;

Query Match 46.5%; Score 59; DB 10; Length 306;  
 Best Local Similarity 47.6%; Pred. No. 9.1;  
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAP 21  
 125 PEYQPPPTSSAPPMYSATSV 145

Db

RESULT 7

Q9UHM8 PRELIMINARY; PRT; 433 AA.

AC Q9UHM8

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Suppressor of fused.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung;

RX MEDLINE=20033603; PubMed=10564661;

RA Stone D.M., Murone M., Luoh S., Ye W., Armanini M.P., Gurney A.,

RA Phillips H., Brush J., Goddard A., deSautage F.J., Rosenthal A.;

RT "Characterization of the human suppressor of fused, a negative

RT regulator of the zinc-finger transcription factor Gli.";

RL J. Cell Sci. 112:4437-4448(1999).

DR EMBL; AF144231; AAF23890.1; -.

DR InterPro; IPR007768; SUFU.



DR Pfam; PF05076; SUFU; 1.  
 SQ SEQUENCE 433 AA; 47932 MW; 40CF1CDD5F526102 CRC64;  
 Query Match 46.5%; Score 59; DB 4; Length 433;  
 Best Local Similarity 59.1%; Pred. No. 13;  
 Matches 13; Conservative 2; Mismatches 5; Indels 2; Gaps 1;  
 QY 2 DTRP--APGSTAPPAGHGVTSAP 21  
 : ||| ||| ||| ||| ||| ||| :  
 Db 3 ELRPSGAPGPTAPPAGPTAPP 24  
 : ||| ||| ||| ||| ||| ||| :  
 RESULT 8  
 Q9NZ07  
 ID Q9NZ07 PRELIMINARY; PRT; 481 AA.  
 AC Q9NZ07;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Suppressor of fused variant 3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=20033603; PubMed=10564661;  
 RA Stone D.M., Murone M., Luoh S., Ye W., Armanini M.P., Gurney A.,  
 RA Phillips H., Brush J., Goddard A., deSauvage F.J., Rosenthal A.;  
 RT "Characterization of the human suppressor of fused, a negative  
 RT regulator of the zinc-finger transcription factor Gli.";  
 RL J. Cell Sci. 112:4437-4448(1999).  
 RM [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Stone D.M., Brush J., Goddard A., Rosenthal A.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF222345; AAF35866.1; -.  
 DR GO; GO:0005737; C:cytoplasm; NAS.  
 DR GO; GO:0003714; F:transcription co-repressor activity; TAS.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.  
 DR InterPro; IPR007768; SUFU.  
 DR Pfam; PF05076; SUFU; 1.  
 SQ SEQUENCE 481 AA; 52976 MW; 7B312507FF653923 CRC64;

Query Match 46.5%; Score 59; DB 4; Length 481;  
 Best Local Similarity 59.1%; Pred. No. 14;  
 Matches 13; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY 2 DTRP--APGSTAPPAGHGVTSAP 21  
 : ||| ||| ||| ||| ||| ||| :  
 Db 3 ELRPSGAPGPTAPPAGPTAPP 24  
 : ||| ||| ||| ||| ||| ||| :  
 RESULT 9  
 Q9UMX1  
 ID Q9UMX1 PRELIMINARY; PRT; 484 AA.  
 AC Q9UMX1;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Suppressor of fused.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20032725; PubMed=10559945;  
 RA Kogerman P., Grimm T., Kogerman L., Krause D., Unden A.B.,  
 RA Stastfeldt B., Toftgard R., Zaphiropoulos P.G.;  
 RT "Mammalian suppressor-of-fused modulates nuclear-cytoplasmic shuttling

RT of Gli-1.";  
 RL Nat. Cell Biol. 1:312-319(1999).  
 DR EMBL; AF175770; AAD50501.1; -.  
 DR GO; GO:0005737; C:cytoplasm; TAS.  
 DR GO; GO:0005634; C:nucleus; TAS.  
 DR GO; GO:0004871; F:signal transducer activity; TAS.  
 DR GO; GO:0003714; F:transcription co-repressor activity; TAS.  
 DR GO; GO:0007275; P:development; TAS.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR GO; GO:0001501; P:skeletal development; TAS.  
 DR InterPro; IPR007768; SUFU.  
 DR Pfam; PF05076; SUFU; 1.  
 SQ SEQUENCE 484 AA; 53972 MW; 0B9CC1DE78817AF6 CRC64;

Query Match 46.5%; Score 59; DB 4; Length 484;  
 Best Local Similarity 59.1%; Pred. No. 14;  
 Matches 13; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY 2 DTRP--APGSTAPPAGHGVTSAP 21  
 : ||| ||| ||| ||| ||| ||| :  
 Db 3 ELRPSGAPGPTAPPAGPTAPP 24  
 : ||| ||| ||| ||| ||| ||| :  
 RESULT 10  
 Q9UHK2  
 ID Q9UHK2 PRELIMINARY; PRT; 484 AA.  
 AC Q9UHK2; Q9UMY0;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Suppressor of fused.  
 GN SUFU.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=20033603; PubMed=10564661;  
 RA Stone D.M., Murone M., Luoh S., Ye W., Armanini M.P., Gurney A.,  
 RA Phillips H., Brush J., Goddard A., deSauvage F.J., Rosenthal A.;  
 RT "Characterization of the human suppressor of fused, a negative  
 RT regulator of the zinc-finger transcription factor Gli.";  
 RL J. Cell Sci. 112:4437-4448(1999).  
 RM [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 3-484 FROM N.A.  
 RX MEDLINE=21150643; PubMed=11252182;  
 RA Delattre M., Briand S., Paces-Pessy M., Blanchet-Tournier M.F.;  
 RT "Suppressor of fused gene involved in hedgehog signal transduction in  
 RT Drosophila melanogaster is conserved in mammals.";  
 RL Dev. Genes Evol. 209:294-300(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22084371; PubMed=12068298;  
 RA Taylor M.D., Liu L., Raffel C., Hui C.-C., Mainprize T.G., Zhang X.,  
 RA Agatep R., Chiappa S., Gao L., Lowrance A., Hao A., Goldstein A.M.,  
 RA Stavrou T., Scherer S.W., Dura W.T., Mainwright B., Squire J.A.,  
 RA Rutka J.T., Hogg D.;  
 RT "Mutations in SUFU predispose to medulloblastoma.";  
 RL Nat. Genet. 31:306-310(2002).  
 DR EMBL; AF159447; AAF23893.1; -.  
 DR EMBL; BC013291; AAH13291.1; -.  
 DR EMBL; AF172319; AAD51655.1; -.  
 DR EMBL; AY081829; AAM08947.1; -.  
 DR EMBL; AY081818; AAM08947.1; JOINED.  
 DR EMBL; AY081819; AAM08947.1; JOINED.

```

DR EMBL; AY081820; AAM08947.1; JOINED.
DR EMBL; AY081821; AAM08947.1; JOINED.
DR EMBL; AY081822; AAM08947.1; JOINED.
DR EMBL; AY081823; AAM08947.1; JOINED.
DR EMBL; AY081824; AAM08947.1; JOINED.
DR EMBL; AY081825; AAM08947.1; JOINED.
DR EMBL; AY081826; AAM08947.1; JOINED.
DR EMBL; AY081827; AAM08947.1; JOINED.
DR EMBL; AY081828; AAM08947.1; JOINED.
DR Genew; HGNC:16466; SUFU.
DR InterPro; IPR007768; SUFU.
DR Pfam; PF05076; SUFU.
SQ SEQUENCE 484 AA; 53946 MW; 4A9CDICF75FC179A CRC64;

Query Match
Best Local Similarity 46.5%; Score 59; DB 4; Length 484;
Matches 13; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY 2 DTRP--APGSTAPPAGHGVTSAP 21
Db 3 ELRPSGAPGPTAPPAGPTAPP 24

RESULT 11
Q99109 PRELIMINARY; PRT; 652 AA.
AC Q99109;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Repellent protein.
GN REP1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FB11;
RX MEDLINE=97015128; PubMed=8861956;
RA Wosten H.A., Bohlmann R., Eckerskorn C., Lottspeich F., Bolker M.,
RA Kahmann R.;
RT "A novel class of small amphipathic peptides affect aerial hyphal
RT growth and surface hydrophobicity in Ustilago maydis.";
RL EMBO J. 15:4274-4281(1996).
DR EMBL; U56826; AAC49419.1; -.
DR PIR; S71753; S71753.
SQ SEQUENCE 652 AA; 71373 MW; 07B96277A2E496AB CRC64;

Query Match
Best Local Similarity 46.5%; Score 59; DB 3; Length 652;
Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23
Db 485 PSTKPSPTTPPPSKOTSKPTT 507

RESULT 12
Q19130 PRELIMINARY; PRT; 704 AA.
ID Q19130;
AC Q19130;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gelatinase B.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97467354; PubMed=9325284;

```

```

RA Fang K.C., Raymond W.W., Blount J.L., Caughey G.H.;
RT "Dog mast cell alpha-chymase activates progelatinase B by cleaving the
RT Phe88-Gln89 and Phe91-Glu92 bonds of the catalytic domain.";
RL J. Biol. Chem. 272:25628-25635(1997).
DR EMBL; U89842; AAB81681.1; -.
DR HSP; P08254; 1CIZ.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000562; FN Type II.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR006026; Peptidase M.
DR InterPro; IPR001818; Pept_M10A_M12B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR006970; PT.
DR Pfam; PF00040; fn2; 3.
DR Pfam; PF00045; hemopexin; 4.
DR Pfam; PF00413; Peptidase M10; 1.
DR Pfam; PF03933; Peptidase_M10_N; 1.
DR Pfam; PF04886; PT; 1.
DR PRINTS; PR00013; FNYPEII.
DR PRINTS; PR00133; MATRIXIN.
DR PRODOM; PD000935; FN_Type_II; 3.
DR SMART; SM00059; FN2; 3.
DR SMART; SM00120; HX; 4.
DR SMART; SM00235; ZmC; 1.
DR PROSITE; PS00546; CYSTEINE SWITCH; 1.
DR PROSITE; PS00023; FIBRONECTIN 2; 3.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
KW Hydrolase; Metal-binding; Metalloprotease; Protease; Zinc.
SQ SEQUENCE 704 AA; 78259 MW; AF9999247E463318 CRC64;

Query Match
Best Local Similarity 45.7%; Score 58; DB 6; Length 704;
Matches 11; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAP 21
Db 474 PSERTAGTGPAPAGTGP 494

RESULT 13
Q89X06 PRELIMINARY; PRT; 745 AA.
ID Q89X06;
AC Q89X06;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BIR0521 protein.
GN BIR0521.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005936; BAC45786.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; P:protein kinase activity; IEA.
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR006665; OmpA/MotB.

```

```
DR InterPro; IPR003882; Pistil extensin.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR ProDom; PD000001; Prot_kinase; 1.
KW Complete proteome.
SQ SEQUENCE 745 AA; 74545 MW; 155EDFC74DBC6D CRC64;

Query Match 45.3%; Score 57.5; DB 16; Length 745;
Best Local Similarity 31.2%; Pred. No. 34;
Matches 15; Conservative 1; Mismatches 7; Indels 25; Gaps 1;

QY 1 PDTRPAPGST-----APPAGHGVTSAPDT 23
Db 234 PTATPAPGSTPGAPPAGRPAGPPGVRPGSPPAAGSPAPGATPPT 281

RESULT 14
Q7XLQ4 PRELIMINARY; PRT; 650 AA.
AC Q7XLQ4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNBa0044M19.2 protein.
GN OSJNBa0044M19.2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding G.F.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL731601; CAE05015.1; --
SQ SEQUENCE 650 AA; 69443 MW; 9695477250B4F957 CRC64;

Query Match 44.9%; Score 57; DB 10; Length 650;
Best Local Similarity 58.8%; Pred. No. 35;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 PAPGSTAPPAGHGVTSAP 21
Db 460 PAPGQATPPPGVQYAP 476

RESULT 15
Q9LEY7 PRELIMINARY; PRT; 407 AA.
AC Q9LEY7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN T22D6.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terryn N., Ardiles W., Buysschaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
```

```
RA Villarroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL357612; CAB93721.1; --
DR PIR; T50505; T50505.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
KW Hypothetical protein.
SQ SEQUENCE 407 AA; 44568 MW; B7C0ED222DF2D19B CRC64;

Query Match 44.5%; Score 56.5; DB 10; Length 407;
Best Local Similarity 45.5%; Pred. No. 25;
Matches 10; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22
Db 52 PTSSPPSSSSPPFHG-SNSPD 72

RESULT 16
Q8L708 PRELIMINARY; PRT; 409 AA.
AC Q8L708;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (At5g08200).
GN AT5G08200.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Heuan V.W., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Oncidera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis CRF clones."
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY140035; AAM98176.1; --
DR EMBL; BT008509; AAP37868.1; --
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
KW Hypothetical protein.
SQ SEQUENCE 409 AA; 44846 MW; D2175970E8D8FE40 CRC64;

Query Match 44.5%; Score 56.5; DB 10; Length 409;
Best Local Similarity 45.5%; Pred. No. 25;
Matches 10; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

QY 1 PDTRPAPGSTAPPAGHGVTSAPD 22
Db 54 PTSSPPSSSSPPFHG-SNSPD 74
```

RESULT 17  
Q9RD79 ID Q9RD79 PRELIMINARY; PRT; 411 AA.  
AC Q9RD79;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical protein SC00798.  
GN SC00798 OR SCF43.09.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21936410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL939106; CAB66198.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 411 AA; 41499 MW; BE906AFB7D84926B CRC64;

Query Match 44.1%; Score 56; DB 16; Length 411;  
Best Local Similarity 55.6%; Pred. No. 29;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
QY 5 PAPGSTAPPAHGVTSAPD 22  
DB 200 PGPQGRGPHAHQVVTAPD 217

RESULT 18  
Q9BLR4 ID Q9BLR4 PRELIMINARY; PRT; 508 AA.  
AC Q9BLR4;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN L654.06.  
OS Leishmania major.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Friedlin;  
RA Tosato V., Ciaroni L., Bianchetti G., Bruschi C.V., Ivens A.C.,  
RA Quail M., Rajandream M.A., Barrell B.G.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Friedlin;  
RX MEDLINE=98146435; PubMed=9477341;  
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
RA Smith D.F.;  
RT "A physical map of the Leishmania major Friedlin genome.";  
RL Genome Res. 8:135-145(1998).  
DR EMBL; AL512294; CAC29454.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 508 AA; 54481 MW; ECDBD4173D2A2A48 CRC64;

Query Match 44.1%; Score 56; DB 5; Length 508;  
Best Local Similarity 52.4%; Pred. No. 36;  
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;  
QY 1 PDTRPAPGSTAPPAHGVTSAP 21  
DB 35 PLTAAAPASSCTPPHGTTSPP 55

RESULT 19  
Q7XZD2 ID Q7XZD2 PRELIMINARY; PRT; 544 AA.  
AC Q7XZD2;  
DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Prolyl-tRNA synthetase.  
GN RSPORS1.  
OS Raphanus sativus (Radish).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Raphanus.  
OX NCBI\_TaxID=3726;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hara M., Sugano Y., Kuboi T.;  
RT "Drought-regulated expression of prolyl-tRNA synthetase genes in  
RT radish (Raphanus sativus) seedlings.";  
RL Plant Sci. 165:129-137(2003).  
DR EMBL; AB097406; BAC78195.1; --  
KW Aminoacyl-tRNA synthetase.  
SQ SEQUENCE 544 AA; 60853 MW; 2B35326B36BAC837 CRC64;

Query Match 44.1%; Score 56; DB 10; Length 544;  
Best Local Similarity 43.5%; Pred. No. 39;  
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;  
QY 1 PDTRPAPGSTAPPAHGVTSAPD 23  
DB 32 PRNRPSGFATAPSSGTAASPD 54

RESULT 20  
Q9VWS5 ID Q9VWS5 PRELIMINARY; PRT; 1895 AA.  
AC Q9VWS5;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE CG15040 protein.  
GN CG15040.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weirstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003509; AAF48863.1; -.
DR FlyBase; FBgn0030940; CG15040.
SQ SEQUENCE 1895 AA; 208652 MW; DB8540F2EC4A1CA8 CRC64;

Query Match 44.1%; Score 56; DB 5; Length 1895;
Best Local Similarity 52.4%; Pred.No. 1.4e+02;
Matches 11; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSP 21
Db 596 PAPVPAPAPAPAPAPAPAPAPAP 616

RESULT 21
Q9KGV9
ID Q9KGV9 PRELIMINARY; PRT; 569 AA.
AC Q9KGV9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE P64 protein (lmo2185 protein).
GN LMO2185.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LO28 / Serovar 1/2c;
RX MEDLINE=20461248; PubMed=11004200;
RA Borezee E., Msadek T., Durant L., Berche P.;
RT "Identification in Listeria monocytogenes of MecA, a Homologue of the
RT Bacillus subtilis Competence Regulatory Protein.";
RL J. Bacteriol. 182:5931-5934(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Bojand J.-A., Voss H., Wehland J., Cossart P.;
```

```
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AF282221; AAF82249.1; -.
DR EMBL: AL591982; CAD00263.1; -.
DR PIR: A11347; A11347.
DR Listlist; LMO02185; -.
DR InterPro; IPR006635; NEA_transpt.
DR Pfam; PF05031; NEAT; 3.
DR SMART; SM00725; NEAT; 3.
KW Complete proteome.
SQ SEQUENCE 569 AA; 63380 MW; 702B6382193D2784 CRC64;

Query Match 43.7%; Score 55.5; DB 16; Length 569;
Best Local Similarity 52.2%; Pred.No. 47;
Matches 12; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 1 PDTRPAPGSTAPPAHGVTSP 23
Db 328 PVTKPDGTTNPP---VTTPPTT 347

RESULT 22
Q9ZXJ0
ID Q9ZXJ0 PRELIMINARY; PRT; 130 AA.
AC Q9ZXJ0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Orf35.
OS Bacteriophage phi CTX.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
OX NCBI_TaxID=35343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=phiCTX-c;
RA Hayashi T.;
RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=phiCTX-c;
RX MEDLINE=99157549; PubMed=10027959;
RA Nakayama K., Kanaya S., Ohnishi M., Terawaki Y., Hayashi T.;
RT "The complete nucleotide sequence of phiCTX, a cytotoxin-converting
RT phage of Pseudomonas aeruginosa: implications for phage evolution and
RT horizontal gene transfer via bacteriophage.";
RL Mol. Microbio. 31:399-419(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=phiCTX-c;
RX MEDLINE=90014160; PubMed=2507866;
RA Hayashi T., Kamio Y., Hishinuma F., Usami Y., Titani K., Terawaki Y.;
RT "Pseudomonas aeruginosa cytotoxin; the nucleotide sequence of the gene
RT and the mechanism of activation of the protoxin.";
RL Mol. Microbiol. 3:861-868(1989).
DR EMBL; AB008550; BAA36263.1; -.
SQ SEQUENCE 130 AA; 14356 MW; 3671A0A6924FBE0F CRC64;

Query Match 43.3%; Score 55; DB 9; Length 130;
Best Local Similarity 51.9%; Pred.No. 12;
Matches 14; Conservative 1; Mismatches 6; Indels 6; Gaps 2;

QY 1 PDTRP----APGSTAPPAHGVTSP 23
Db 102 PDGFLSARAPAPQAPPAHPV--IPET 126

RESULT 23
Q8SQ36
ID Q8SQ36 PRELIMINARY; PRT; 160 AA.
AC Q8SQ36;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```





AC Q9SM15;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE SBP-domain protein 5 (Fragment).  
GN SBP5.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. T232; TISSUE=Flower;  
RX MEDLINE=97446501; PubMed=9301089;  
RA Cardon G.H., Hoehmann S., Nettesheim K., Saedler H., Huijser P.;  
RT "Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a  
RT novel gene involved in the floral transition.";  
RL Plant J. 12:367-377(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. T232; TISSUE=Flower;  
RX MEDLINE=99453765; PubMed=10524240;  
RA Cardon G.H., Hoehmann S., Klein J., Nettesheim K., Saedler H.,  
RA Huijser P.;  
RT "Molecular characterization of the Arabidopsis SBP-box genes.";  
RL Gene 237:91-104(1999).  
DR EMBL; AJ011618; CAB56631.1; -.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR004333; SBP.  
DR Pfam; PF03110; SBP; 1.  
DR NON TER 1  
FT SEQUENCE 329 AA; 34775 MW; C344BE2583C8FD83 CRC64;  
SQ  
  
Query Match 43.3%; Score 55; DB 10; Length 329;  
Best Local Similarity 90.9%; Pred. No. 31;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 RPAPGSTAPPA 14  
| | | | | | | | | |  
Db 297 RPAPGSAAPPA 307  
  
RESULT 28  
Q95L89 PRELIMINARY; PRT; 554 AA.  
AC Q95L89;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Mucin (Fragment).  
GN MUC1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sacchi P., Rasero R., Rosati S.;  
RT "Molecular analysis of MUC1 polymorphism in cattle.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF399757; AAL28023.1; -.  
DR InterPro; IPR002965; P rich extensn.  
DR InterPro; IPR000082; SEA\_domain.  
DR Pfam; PF01390; SEA; 1.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00200; SEA; 1.  
DR PROSITE; PS50024; SEA; 1.  
DR NON TER 554  
FT SEQUENCE 554 AA; 55508 MW; E441CD140093AFFD CRC64;  
SQ

Query Match 43.3%; Score 55; DB 6; Length 554;  
Best Local Similarity 47.6%; Pred. No. 53;  
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
  
QY 1 PDTRPAPGSTAPPAHGVTSP 21  
| : | | | | | | | | | |  
Db 146 PTSSPAPSPAASPGHGTSSP 166  
  
RESULT 29  
Q8WML4 PRELIMINARY; PRT; 580 AA.  
AC Q8WML4;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE MUC1 protein precursor.  
GN MUC1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Holstein Dairy cow; TISSUE=Lactating mammary gland;  
RX MEDLINE=21672491; PubMed=11814015;  
RA Pallesen L.T., Andersen M.H., Nielsen R.L., Berglund L.,  
RA Rasmussen L.K., Petersen T.E., Rasmussen J.T.;  
RT "Purification of MUC1 from bovine milk-fat globules and  
RT characterization of a corresponding full-length cDNA Clone.";  
RL J. Dairy Sci. 84:2591-2598(2001).  
DR EMBL; AJ400824; CAC81810.1; -.  
DR InterPro; IPR002965; P rich extensn.  
DR InterPro; IPR000082; SEA\_domain.  
DR Pfam; PF01390; SEA; 1.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00200; SEA; 1.  
DR PROSITE; PS50024; SEA; 1.  
KW Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 580 MUC1 PROTEIN.  
SQ SEQUENCE 580 AA; 58091 MW; E91C13984AF7D757 CRC64;  
  
Query Match 43.3%; Score 55; DB 6; Length 580;  
Best Local Similarity 47.6%; Pred. No. 56;  
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
  
QY 1 PDTRPAPGSTAPPAHGVTSP 21  
| : | | | | | | | | | |  
Db 146 PTSSPAPSPAASPGHGTSSP 166  
  
RESULT 30  
O19115 PRELIMINARY; PRT; 602 AA.  
AC O19115;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Muc1 (Fragment).  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE OF 410-602 FROM N.A.  
RC STRAIN=New Zealand White;  
RX MEDLINE=96351712; PubMed=8747930;  
RA Spicer A.P., Iuhig T., Chilton B.S., Gendler S.J.;  
RT "Analysis of mammalian MUC1 genes reveals potential functionally  
RT important domains.";  
RL Mamm. Genome 6:885-888(1995).  
RL

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand White;
RA Hewetson A., Chilton B.S.;
RT "Molecular cloning and hormone-dependent expression of rabbit Muc1 in
RT the cervix and uterus.";
RL Biol. Reprod. 0:0-0(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand White;
RA Hewetson A., Chilton B.S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85787; AAB64380.1; -.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SMO0200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON TER 1
SQ SEQUENCE 602 AA; 61287 MW; 0F4523CF2871F270 CRC64;

Query Match 43.3%; Score 55; DB 6; Length 602;
Best Local Similarity 52.2%; Pred. No. 58;
Matches 12; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 PDTRPAGSTAPPAGHGVTSAPDT 23
Db ||||| ||||| ||||| ||||| |||||
187 PATSPTSVSATSPVHEVTSAPAT 209

RESULT 31
O45360 PRELIMINARY; PRT; 916 AA.
AC O45360; O45570;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE F14F7.5 protein.
GN F14F7.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RA McMurray A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z81548; CAB04465.1; -.
DR EMBL; Z81503; CAB04465.1; JOINED.
DR EMBL; Z81503; CAB04114.1; -.
DR EMBL; Z81548; CAB04114.1; JOINED.
DR PIR; T20909; T20909.
DR WormPep; F14F7.5; CE15825.
DR InterPro; IPR003125; WSN.

```

```

DR Pfam; PF02206; WSN; 1.
DR SMART; SM00453; WSN; 1.
SQ SEQUENCE 916 AA; 101775 MW; A2950E712D147501 CRC64;

Query Match 43.3%; Score 55; DB 5; Length 916;
Best Local Similarity 44.0%; Pred. No. 88;
Matches 11; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

QY 1 PDTRPAPG--STAPPAGHGVTSAPDT 23
Db ||||| ||||| ||||| ||||| |||||
796 PPSQPCPGPRGSSPPAPSAPSAPET 820

RESULT 32
Q9ULK3 PRELIMINARY; PRT; 1339 AA.
AC Q9ULK3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA1217 (Fragment).
GN KIAA1217.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirotsawa M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DE EMBL; AB033043; BAA86531.2; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 1339 AA; 146176 MW; 474485C5314E8A56 CRC64;

Query Match 43.3%; Score 55; DB 4; Length 1339;
Best Local Similarity 52.9%; Pred. No. 1.3e+02;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 PAPGSTAPPAGHGVTSAP 21
Db ||||| ||||| ||||| ||||| |||||
293 PRPGSTAHPHPAIPNSP 309

RESULT 33
Q8IQ93 PRELIMINARY; PRT; 509 AA.
AC Q8IQ93;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CG8042-PB.
GN BCNA:GH10229 OR CG8042.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003557; AAN12033.1; -.  
DR FlyBase; FBgn0027554; BcDNA:GH10229.  
DR InterPro; IPR001012; UBX.  
DR Pfam; PF00789; UBX; 1.  
DR SMART; SM00166; UBX; 1.  
DR PROSITE; PS50033; UBX; 1.

SQ SEQUENCE 509 AA; 55687 MW; F3C0094E3C403B0B CRC64;

Query Match 42.9%; Score 54.5; DB 5; Length 509;  
Best Local Similarity 50.0%; Pred. No. 56;  
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 PDTRPAPGSTAPPAGVTSQA 20  
| : ||| : ||| : |  
Db 440 PQNQPQPGNT-PPAHTSQA 458

RESULT 34

QYI14 PRELIMINARY; PRT; 656 AA.

AC Q9YI14;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE BCDNA:GH10229 protein.  
GN BCDNA:GH10229 OR CG8042.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;

[1]  
RN SEQUENCE FROM N.A.

RP STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Easus A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,  
RA Aghayani A., Arcaina T.T., Baxter E., Blazej R.G., Butenhoff C.,  
RA Champagne M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,  
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,

```
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,
RA Park S., Sequeira A., Sethi H., Snir E., Svirskas R.R., Weinburg T.,
RA Celniker S.E.;
RA "Full length Drosophila melanogaster cDNA sequence.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003557; AAF50501.1; -.
DR EMBL; AF145658; AAD38633.1; -.
DR FlyBase; FBgn027554; BcDNA:GH10229.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR006663; Thioresox_dom2.
DR InterPro; IPR001012; UBX.
DR Pfam; PF00789; UBX; 1.
DR SMART; SM00166; UBX; 1.
DR PROSITE; PS50033; UBX; 1.
SQ SEQUENCE 656 AA; 71204 MW; 6374A7D2608CACIB CRC64;

Query Match 42.9%; Score 54.5; DB 5; Length 656;
Best Local Similarity 50.0%; Pred. No. 73;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 PDTRPAPGSTAPPAHGVTSAP 20
Db 587 PQNQPQPGGNT-PPAHNTSQA 605

RESULT 35
Q8TE25 PRELIMINARY; PRT; 85 AA.
AC Q8TE25;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Maltase-glucoamylase (Fragment).
GN MGAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nichols B.L. Jr., Avery S.E., Sen P., Swallow D.M., Hahn D.,
RA Sterchi E.;
RT "Sequencing The Maltase-Glucoamylase Gene: Common Ancestry to Sucrase-
RT Isomaltase With Complementary Starch Digestion.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF432186; AAL83559.1; -.
DR EMBL; AF432184; AAL83559.1; JOINED.
FT NON_TER 1
FT NON_TER 85
SQ SEQUENCE 85 AA; 8533 MW; 5B99EC806A042B5B CRC64;

Query Match 42.5%; Score 54; DB 4; Length 85;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 3 TRPAPGSTAPPAHGVTSAPD 22
Db 31 TAPDPGTTGTPDPGTTGTPD 50

RESULT 36
Q00881 PRELIMINARY; PRT; 455 AA.
AC Q00881;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Negative acting factor (Fragment).
GN NAF.
OS Nectria haematococca.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Nectria.
```

```
OX NCBI_TaxID=140110;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=T-8;
RA Li D., Kolattukudy P.E.;
RT "A DNA-binding protein containing a C-terminal Cys62n2 binuclear
RT cluster motif may be involved in regulating the cutinase gene.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR CLUSTER
CC DOMAIN.
DR EMBL; U62028; AAB05250.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR001138; Fungi_Trn.
DR Pfam; PF00172; Zn_c1us; 1.
DR SMART; SM00066; GAL4; 1.
DR PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR PROSITE; PS50048; ZN2_CY6_FUNGAL_2; 1.
KW DNA-binding; Metal-binding; Nuclear protein; Transcription;
KW Transcription regulation; Zinc.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 455 AA; 46383 MW; 471FA1DD8DCD632 CRC64;

Query Match 42.5%; Score 54; DB 3; Length 455;
Best Local Similarity 47.6%; Pred. No. 58;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAPPAHGVTSAP 21
Db 17 PSLAPTAASAAPPAYAAASAP 37

RESULT 37
Q8CD55 PRELIMINARY; PRT; 508 AA.
AC Q8CD55;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Zyxin.
GN ZYX OR 9530098H06RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK031415; BAC27394.1; -.
DR MGD; MGI:103072; Zyx.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 3.
DR ProDom; PD000094; LIM; 3.
DR SMART; SM00132; LIM; 3.
DR PROSITE; PS00150; CYTOCHROME_C; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 2.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
SQ SEQUENCE 508 AA; 54800 MW; 27802B255595632E CRC64;

Query Match 42.5%; Score 54; DB 11; Length 508;
```



Best Local Similarity 75.0%; Pred. No. 65;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAP 12  
|:|||||  
Db 130 PSTKPAPGGTAP 141

## RESULT 38

Q7TOE2  
ID Q7TOE2 PRELIMINARY; PRT; 533 AA.  
AC Q7TOE2;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10093;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Altshuler S.D., Collins E.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC054775; AHH54775.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 533 AA; 57027 MW; 1396D0A256AC12CD CRC64;  
Query Match 42.5%; Score 54; DB 11; Length 533;  
Best Local Similarity 75.0%; Pred. No. 68;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PDTRPAPGSTAP 12  
|:|||||  
Db 155 PSTKPAPGGTAP 166

## RESULT 39

Q9XA04  
ID Q9XA04 PRELIMINARY; PRT; 576 AA.  
AC Q9XA04;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative serine/threonine protein kinase.  
GN SC03860 OR SCH69.30.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1302;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Murphy L., Harris D.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AL939118; CAB45227.1; -.  
DR PIR; T36729; T36729.  
DR GO; GO:0005524; P:ATP binding; IEA.  
DR GO; GO:0004674; P:protein serine/threonine kinase activity; IEA.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0006463; P:protein amino acid phosphorylation; IEA.  
DR InterPro; IPR000719; Prot kinase.  
DR InterPro; IPR008271; Ser\_thr\_kin\_AS.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot kinase; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase;  
Complete proteome.  
SQ SEQUENCE 576 AA; 59816 MW; 3A8DE0A5E70B8C69 CRC64;

Query Match 42.5%; Score 54; DB 16; Length 576;  
Best Local Similarity 55.6%; Pred. No. 74;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 RPAPGSTAPPAHGVTSAP 21  
|:|||||  
Db 347 RPSGSGPPPTGPDSTP 364

## RESULT 40

Q8NLV6  
ID Q8NLV6 PRELIMINARY; PRT; 635 AA.  
AC Q8NLV6;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical membrane protein Cgl2830.  
GN CGL2830.  
OS Corynebacterium glutamicum (Brevibacterium flavum).

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RA Nakagawa S.;  
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005283; BAC00224.1; -.  
 DR InterPro; IPR002965; P rich extensn.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 635 AA; 66081 MW; ADFC4B8CE943C4EE CRC64;  
  
 Query Match 42.5%; Score 54; DB 16; Length 635;  
 Best Local Similarity 41.4%; Pred. No. 82;  
 Matches 12; Conservative 2; Mismatches 5; Indels 10; Gaps 1;  
  
 Qy 5 PAPGSTAPPAGHVT-----SAPDT 23  
 ||||| ||||| :  
 Db 198 PVPGSVTPPAPGISAPGGALFTPGSAPPT 226

Search completed: May 6, 2004, 16:31:07  
 Job time : 44.0513 secs